ELSEVIER

Journal of Econometrics 90 (1999) 77–97

JOURNAL OF Econometrics

Distribution-free estimation of some nonlinear panel data models

Jeffrey M. Wooldridge*

Department of Economics, Michigan State University, East Lansing, MI 48824-1038, USA

Received 1 October 1996; received in revised form 1 January 1998

Abstract

This paper studies distribution-free estimation of some multiplicative unobserved components panel data models. One class of estimators requires only specification of the conditional mean; in particular, the multinomial quasi-conditional maximum likelihood estimator is shown to be consistent when only the conditional mean in the unobserved effects model is correctly specified. Additional orthogonality conditions can be used in a method of moments framework. A second class of problems specifies the conditional mean, conditional variances, and conditional covariances. Using the notion of a conditional linear predictor, it is shown that specification of conditional second moments implies further orthogonality conditions in the observable data that can be exploited for efficiency gains. This has applications to both count and gamma-type panel data regression models. © 1999 Elsevier Science S.A. All rights reserved.

JEL classification: C23

Keywords: Fixed effects Poisson model; Panel data; Robustness; Quasi-conditional maximum likelihood; Generalized method of moments

1. Introduction

In the standard linear unobserved effects model, popular estimators are consistent under correct specification of the conditional mean and strict exogeneity of the explanatory variables, conditional on the latent individual

* E-mail: wooldri1@pilot.msu.edu

PII: S0304-4076(98)00033-5

effect. The usual fixed effects (within) estimator is consistent, as is the minimum chi-square estimator proposed by Chamberlain (1982).

Less has been written on distribution-free estimation of nonlinear unobserved effects models. Most of the work on nonlinear panel data methods relies on the method of conditional maximum likelihood (CML), where a sufficient statistic (the sum of the explained variable across time) is conditioned on to remove the unobserved effect. Examples include Chamberlain (1980,1984) for binary responses and Hausman et al. (1984) (hereafter HHG) for count data. The robustness properties of these CMLEs to misspecification of the initially specified joint distribution have not been investigated. It is important to stress that, even though the resulting conditional density (e.g., the multinomial) is typically in the linear exponential family (LEF), the robustness of the quasi-CMLE (QCMLE) in the LEF, e.g., Gourieroux et al. (1984) (hereafter GMT), cannot be appealed to. This is because, except in special cases, the expectation associated with the LEF conditional density is misspecified if the initial joint distribution in the unobserved components model is misspecified.

For models of nonnegative variables, we would like to have a class of estimators that requires minimal distributional assumptions while further relaxing the first two moment assumptions appearing in the literature. In independent work, Chamberlain (1992a) has offered a method of moments approach to nonlinear unobserved components models that applies to multiplicative models as a special case. All that is the required is the specification of the mean conditional on the exogenous variables and the latent effects.

This paper reports on results first derived in Wooldridge (1990b), and derives consistent estimators of multiplicative unobserved effects models under two kinds of assumptions. The first class of models specifies only the conditional mean, as in Chamberlain (1992a). However, I offer a direct proof that the multinomial QCMLE, also known as the fixed effects Poisson (FEP) estimator, consistently estimates the conditional mean parameters, and this leads naturally into method of moments estimators that could improve on the efficiency of the FEP estimator. Second, I consider models that specify the conditional variances and covariances in addition to the conditional means. For a count model with individual-specific dispersion and an unobserved effects gamma-type regression model, further orthogonality conditions are implied that can be exploited in estimation.

Section 2 introduces the count model that motivated this research, and discusses the assumptions that have been imposed in previous work. Section 3 discusses the estimation of models that specify only the conditional mean, and shows that the fixed effects Poisson estimator is consistent very generally. Conditional linear predictors are introduced in Section 4. Section 5 analyzes a count-type model, where the conditional variance is proportional to the conditional mean, and the conditional covariances are zero. Section 6 covers the case where the variance is proportional to the square of the mean, and models with parametric serial correlation are discussed briefly in Section 7.

2. Motivation: An unobserved effects model for count panel data

Let $\{(\boldsymbol{y}_i, \boldsymbol{x}_i, \phi_i): i = 1, 2, \ldots\}$ be a sequence of i.i.d. random variables, where $\boldsymbol{y}_i \equiv (y_{i1}, \ldots, y_{iT})'$ is an observable $T \times 1$ vector of counts, $\boldsymbol{x}_i \equiv (\boldsymbol{x}'_{i1}, \boldsymbol{x}'_{i2}, \ldots, \boldsymbol{x}'_{iT})'$ is a $T \times K$ matrix of observable conditioning variables ($\boldsymbol{x}_{it}$ is $1 \times K$, $t = 1, \ldots, T$), and ϕ_i is an unobservable random scalar. The fixed effects Poisson model analyzed by HHG assumes that

$$y_{it}|\boldsymbol{x}_i, \phi_i \sim \text{Poisson}(\phi_i \mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0)), \quad t = 1, \ldots, T \tag{2.1}$$

and

$$y_{it}, y_{ir} \text{ are independent conditional on } \boldsymbol{x}_i, \phi_i, \quad t \neq r, \tag{2.2}$$

where

$$\mathrm{E}(y_{it}|\boldsymbol{x}_i, \phi_i) = \mathrm{E}(y_{it}|\boldsymbol{x}_{it}, \phi_i) = \phi_i \mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0), \tag{2.3}$$

and $\boldsymbol{\beta}_0$ is a $P \times 1$ vector of unknown parameters. Actually, HHG take $\mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}) \equiv \exp(\boldsymbol{x}_{it}\boldsymbol{\beta})$, but there is no need to use this particular functional form. It is convenient to choose μ so that $\mu(\boldsymbol{x}_{it}, \boldsymbol{\beta})$ is well-defined and positive for all $\boldsymbol{x}_{it}$ and $\boldsymbol{\beta}$. Assumptions (2.1) and (2.2) incorporate strict exogeneity of $\boldsymbol{x}_{it}$ conditional on ϕ_i, independence of y_{it} and y_{ir} conditional on $\boldsymbol{x}_i$ and ϕ_i, and the Poisson distributional assumption.

HHG use Andersen's (1970) conditional ML methodology to estimate $\boldsymbol{\beta}_0$ under Eqs. (2.1) and (2.2). Let $n_i \equiv \sum_{t=1}^{T} y_{it}$ denote the sum across time of the explained variable. Then HHG show (see also Palmgren, 1981) that

$$\boldsymbol{y}_i|n_i, \boldsymbol{x}_i, \phi_i \sim \text{Multinomial}(n_i, p_1(\boldsymbol{x}_i, \boldsymbol{\beta}_0), \ldots, p_T(\boldsymbol{x}_i, \boldsymbol{\beta}_0)), \tag{2.4}$$

where

$$p_t(\boldsymbol{x}_i, \boldsymbol{\beta}_0) \equiv \mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0) \bigg/ \left(\sum_{r=1}^{T} \mu(\boldsymbol{x}_{ir}, \boldsymbol{\beta}_0) \right). \tag{2.5}$$

Because this distribution does not depend on ϕ_i, Eq. (2.4) is also the distribution of $\boldsymbol{y}_i$ conditional on n_i and $\boldsymbol{x}_i$. Therefore, $\boldsymbol{\beta}_0$ can be estimated by standard conditional ML techniques. For later use, the conditional log-likelihood for observation i, apart from terms not depending on $\boldsymbol{\beta}$, is

$$\ell_i(\boldsymbol{\beta}) = \sum_{t=1}^{T} y_{it} \log[p_t(\boldsymbol{x}_i, \boldsymbol{\beta})]. \tag{2.6}$$

Because the multinomial distribution is in the LEF, the results of GMT imply a certain amount of robustness of the QCMLE. *If*

$$\mathrm{E}(y_{it}|n_i, \boldsymbol{x}_i) = p_t(\boldsymbol{x}_i, \boldsymbol{\beta}_0)n_i \tag{2.7}$$

then the multinomial QCMLE is consistent and asymptotically normal, even if the multinomial distribution is misspecified. Other than the FEP model (2.1) and (2.2), there is another interesting case where Eq. (2.7) holds: HHG's fixed effects negative binomial (FENB) model. Let $\delta_i > 0$ be the unobserved effect. If

$$y_{it}|\boldsymbol{x}_i,\, \delta_i \sim \text{ Negative Binomial}(\mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0),\delta_i) \tag{2.8}$$

(parameterized as in HHG so that $\mathrm{E}(y_{it}|\boldsymbol{x}_i, \delta_i) = \mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0)/\delta_i$) and

$$y_{it}, y_{ir} \text{ are independent conditional on } (\boldsymbol{x}_i, \delta_i), \quad t \neq r, \tag{2.9}$$

then Eq. (2.7) can be shown to hold. (This is asserted in HHG, p. 935, based on features of the negative hypergeometric distribution, which is not quite general enough; in fact, y_{it} given $(n_i, \boldsymbol{x}_i, \delta_i)$ has a binomial-beta distribution with parameters $(\mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0), \sum_{r=1}^{T}\mu(\boldsymbol{x}_{ir}, \boldsymbol{\beta}_0), n_i)$. Then Eq. (2.7) follows from Johnson and Kotz (1969), p. 231). A complete proof is available on request from the author.) By the results of GMT, the QCMLE based on the multinomial distribution provides consistent estimates of $\boldsymbol{\beta}_0$ under Eqs. (2.8) and (2.9). This is a useful result but, as shown in the next section, is much more restrictive than necessary.

A robust approach consists of specifying at most a couple of low-order conditional moments. Chamberlain (1992a) offers a transformation that eliminates ϕ_i, resulting in orthogonality conditions that allow estimation of $\boldsymbol{\beta}_0$ by the method of moments. The next section takes a different approach to this problem, starting by showing that the multinomial QCMLE is consistent and asymptotically normal under just the conditional mean assumption (2.3). This analysis leads to additional orthogonality conditions that can improve efficiency if the full set of assumptions listed in Eqs. (2.1) and (2.2) do not hold. Section 5 examines a model that is more general than Eqs. (2.1) and (2.2) in the first two moments, and suggests methods of estimation.

3. Estimation under conditional mean assumptions

3.1. The quasi-conditional maximum likelihood estimator

Let $\{(\boldsymbol{y}_i, \boldsymbol{x}_i, \phi_i): i = 1, 2, \ldots\}$ be a sequence of i.i.d. random variables, where $\boldsymbol{y}_i$ is $T \times 1$, $\boldsymbol{x}_i$ is $T \times K$, and ϕ_i is an unobserved scalar effect. Consider the model

$$\mathrm{E}(y_{it}|\boldsymbol{x}_i, \phi_i) = \phi_i\mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0), \quad t = 1, \ldots, T, \tag{3.1}$$

where $\boldsymbol{\beta}_0$ is a $P \times 1$ unknown vector. Suppose that $\boldsymbol{\beta}_0$ is estimated by maximizing the log-likelihood associated with the multinomial density, given for each i by Eq. (2.6). It is important to see that this does not require $\boldsymbol{y}_i$ to be a vector of counts. For example, y_{it} could be a binary variable – in which case $0 < \phi_i < 1$ and $\mu(\boldsymbol{x}_{it},\boldsymbol{\beta})$ could be the logit or probit response probability – or a nonnegative continuous variable – in which case a popular choice of $\mu(\boldsymbol{x}_{it}, \boldsymbol{\beta})$ is the exponential function.

We now show that the multinomial QCMLE is consistent for $\boldsymbol{\beta}_0$ under Eq. (3.1) only. There are two ways to establish the consistency of the QCMLE. The first is to show that $\mathrm{E}[\ell_i(\boldsymbol{\beta})]$ is maximized at $\boldsymbol{\beta}_0$, for which we use the following lemma.

Lemma 3.1. Let $y_1, y_2, \ldots, y_T$ be nonnegative random variables with finite, nonzero means $\mu_1^0, \mu_2^0, \ldots, \mu_T^0$, and let $\mathbb{R}^T_{++}$ denote the subset of T-dimensional Euclidean space with strictly positive elements. Then $\mu_0 \equiv (\mu_1^0, \mu_2^0, \ldots, \mu_T^0)$ is the unique solution to

$$\max_{\boldsymbol{\mu} \in \mathbb{R}^T_{++}} \sum_{t=1}^{T} \mu_t^0 \log\left(\mu_t \middle/ \left(\sum_{r=1}^{T} \mu_r\right)\right). \tag{3.2}$$

Proof. Dividing Eq. (3.2) by $\sum_{r=1}^{T} \mu_r^0$ does not change the problem. Thus, write $p_t^0 = \mu_t^0 / (\sum_{r=1}^{T} \mu_r^0)$ and $p_t = \mu_t / (\sum_{r=1}^{T} \mu_r)$. Then, it suffices to show that

$$\sum_{t=1}^{T} p_t^0 \log(p_t^0) > \sum_{t=1}^{T} p_t^0 \log(p_t), \tag{3.3}$$

for all $\boldsymbol{p} = (p_1, \ldots, p_T)' \neq \boldsymbol{p}^0 = (p_1^0, \ldots, p_T^0)'$. But this follows immediately from the Kullback–Leibler information inequality (see, for example, Rao, 1973, p. 59), because the left-hand side is the expected value of the log density that puts mass p_t^0 at point t, and the right-hand side is the expected value of the log of another candidate density over the points $\{1, 2, \ldots, T\}$. $\square$

How does this lemma apply to the QCMLE estimator? From Eq. (2.6) we have $\mathrm{E}[\ell_i(\boldsymbol{\beta})|\boldsymbol{x}_i, \phi_i] = \phi_i \sum_{t=1}^{T} \mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0) \log[p_t(\boldsymbol{x}_i, \boldsymbol{\beta})]$, and so, by Lemma 3.1, $\boldsymbol{\beta}_0$ maximizes $\mathrm{E}[\ell_i(\boldsymbol{\beta})|\boldsymbol{x}_i, \phi_i]$ for any $(\boldsymbol{x}_i, \phi_i)$. The law of iterated expectations then shows that $\boldsymbol{\beta}_0$ maximizes $\mathrm{E}[\ell_i(\boldsymbol{\beta})]$ over the parameter space. Consistency then follows under a standard identification assumption and regularity conditions that ensure the uniform weak law of large numbers holds (see, for example, Newey and McFadden, 1994, Section 2).

A second approach is to show that the score of the QCMLE log-likelihood has a zero expectation when evaluated at $\boldsymbol{\beta}_0$. Because we will need this score for inference and for motivating additional orthogonality conditions, we sketch the derivation:

$$
\begin{aligned}
\boldsymbol{s}_i(\boldsymbol{\beta}) &= \nabla_\beta \ell_i(\boldsymbol{\beta}) = \sum_{t=1}^{T} y_{it}[\nabla_\beta p_t(\boldsymbol{x}_i, \boldsymbol{\beta})'/p_t(\boldsymbol{x}_i, \boldsymbol{\beta})] \\
&= \sum_{t=1}^{T} [\nabla_\beta p_t(\boldsymbol{x}_i, \boldsymbol{\beta})'/p_t(\boldsymbol{x}_i, \boldsymbol{\beta})]\{y_{it} - p_t(\boldsymbol{x}_i, \boldsymbol{\beta})n_i\} \qquad (3.4) \\
&\equiv \nabla_\beta \boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta})' \boldsymbol{W}(\boldsymbol{x}_i, \boldsymbol{\beta})\{\boldsymbol{y}_i - \boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta})n_i\} \\
&\equiv \nabla_\beta \boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta})' \boldsymbol{W}(\boldsymbol{x}_i, \boldsymbol{\beta})\boldsymbol{u}_i(\boldsymbol{\beta}), \qquad (3.5)
\end{aligned}
$$

where $\boldsymbol{W}(\boldsymbol{x}_i, \boldsymbol{\beta}) \equiv [\mathrm{diag}\{p_1(\boldsymbol{x}_i, \boldsymbol{\beta}), \ldots, p_T(\boldsymbol{x}_i, \boldsymbol{\beta})\}]^{-1}$, $\boldsymbol{u}_i(\boldsymbol{\beta}) \equiv \boldsymbol{y}_i - \boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta})n_i$, $\boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta}) \equiv [p_1(\boldsymbol{x}_i, \boldsymbol{\beta}), \ldots, p_T(\boldsymbol{x}_i, \boldsymbol{\beta})]'$, and $p_t(\boldsymbol{x}_i, \boldsymbol{\beta})$ is given by Eq. (2.5). Eq. (3.4) follows from the fact that $\sum_{t=1}^{T} p_t(\boldsymbol{x}_i, \boldsymbol{\beta}) = 1$ for all $\boldsymbol{\beta}$, so that $\sum_{t=1}^{T} \nabla_\beta p_t(\boldsymbol{x}_i, \boldsymbol{\beta}) = 0$ for all $\boldsymbol{\beta}$.

The expectation of $\boldsymbol{s}_i(\boldsymbol{\beta}_0)$ conditional on $\boldsymbol{x}_i$ can be found by computing $\mathrm{E}[\boldsymbol{u}_i(\boldsymbol{\beta}_0)|\boldsymbol{x}_i]$. Let $\boldsymbol{\mu}(\boldsymbol{x}_i, \boldsymbol{\beta}) \equiv [\mu(\boldsymbol{x}_{i1}, \boldsymbol{\beta}), \ldots, \mu(\boldsymbol{x}_{iT}, \boldsymbol{\beta})]'$ be the $T \times 1$ vector of conditional means. Then

$$
\begin{aligned}
\mathrm{E}[\boldsymbol{u}_i(\boldsymbol{\beta}_0)|\boldsymbol{x}_i, \phi_i] &= \mathrm{E}(\boldsymbol{y}_i|\boldsymbol{x}_i, \phi_i) - \boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)\mathrm{E}(n_i|\boldsymbol{x}_i, \phi_i) \\
&= \phi_i \boldsymbol{\mu}(\boldsymbol{x}_i, \boldsymbol{\beta}_0) - \left[\boldsymbol{\mu}(\boldsymbol{x}_i, \boldsymbol{\beta}_0) \Big/ \sum_{t=1}^{T} \mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0) \right] \left[\phi_i \sum_{t=1}^{T} \mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0) \right] \\
&= \phi_i \boldsymbol{\mu}(\boldsymbol{x}_i, \boldsymbol{\beta}_0) - \phi_i \boldsymbol{\mu}(\boldsymbol{x}_i, \boldsymbol{\beta}_0) = \boldsymbol{0}. \qquad (3.6)
\end{aligned}
$$

Thus, $\mathrm{E}[\boldsymbol{u}_i(\boldsymbol{\beta}_0)|\boldsymbol{x}_i] = \boldsymbol{0}$ and therefore $\mathrm{E}[\boldsymbol{s}_i(\boldsymbol{\beta}_0)|\boldsymbol{x}_i] = \boldsymbol{0}$; by iterated expectations,

$$
\mathrm{E}[\boldsymbol{s}_i(\boldsymbol{\beta}_0)] = \boldsymbol{0}, \qquad (3.7)
$$

so that the score evaluated at $\boldsymbol{\beta}_0$ has zero mean under Eq. (3.1) only.

We have effectively shown that the multinomial QCMLE – which is derived from a nominal Poisson assumption – is generally consistent for $\boldsymbol{\beta}_0$ provided only that the conditional mean assumption (3.1) holds. We do not need to assume Eq. (2.4) or make any assumptions about the temporal dependence of the y_{it}. This is a useful result, as multinomial QCMLE is a popular method of estimating fixed effects count models. It also shows that the multinomial QCMLE can be used to obtain consistent estimates of $\boldsymbol{\beta}_0$ under Eq. (3.1) whether or not $\boldsymbol{y}_i$ is a vector of counts. The response y_{it} could be a binary variable, a proportion, a nonnegative continuously distributed random variable, or could have discrete and continuous characteristics: its distribution is not restricted, nor is its temporal dependence.

Unless assumptions (2.1) and (2.2) hold, the usual inverse of the estimated information matrix is not a valid estimator of the asymptotic variance of $\hat{\beta}$.

Generally,

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \overset{\text{a}}{\sim} N(\boldsymbol{0}, \boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1}), \tag{3.8}$$

where $\boldsymbol{A}_0$ is the expected value of the Hessian of $\ell_i(\boldsymbol{\beta}_0)$ and $\boldsymbol{B}_0$ is the expected value of $\boldsymbol{s}_i(\boldsymbol{\beta}_0)\boldsymbol{s}_i(\boldsymbol{\beta}_0)'$. A consistent estimator of $\boldsymbol{A}_0$ (as $N \to \infty$ with T fixed) is easily seen to be

$$\hat{\boldsymbol{A}} \equiv \left(N^{-1} \sum_{i=1}^{N} n_i \nabla_\beta \boldsymbol{p}(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}})' \boldsymbol{W}(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}}) \nabla_\beta \boldsymbol{p}(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}}) \right) \equiv \left(N^{-1} \sum_{i=1}^{N} \nabla_\beta \hat{\boldsymbol{m}}_i' \hat{\boldsymbol{V}}_i^{-1} \nabla_\beta \hat{\boldsymbol{m}}_i \right), \tag{3.9}$$

where $\hat{\boldsymbol{m}}_i \equiv n_i \boldsymbol{p}(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}})$ and $\hat{\boldsymbol{V}}_i \equiv \boldsymbol{V}(n_i, \boldsymbol{x}_i, \hat{\boldsymbol{\beta}}) \equiv \text{diag}\{p_{i1}(\hat{\boldsymbol{\beta}})n_i, \dots, p_{iT}(\hat{\boldsymbol{\beta}})n_i\}$. When Eq. (3.9) is multiplied by N and inverted, we obtain the usual maximum likelihood estimate of $\text{Avar}(\hat{\boldsymbol{\beta}})$, which is familiar from standard likelihood theory involving the multinomial distribution.

A consistent estimator of $\boldsymbol{B}_0$ is

$$\begin{aligned}\hat{\boldsymbol{B}} &\equiv N^{-1} \sum_{i=1}^{N} \boldsymbol{s}_i(\hat{\boldsymbol{\beta}})\boldsymbol{s}_i(\hat{\boldsymbol{\beta}})' = N^{-1} \sum_{i=1}^{N} \nabla_\beta \boldsymbol{p}(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}})' \boldsymbol{W}(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}}) \hat{\boldsymbol{u}}_i \hat{\boldsymbol{u}}_i' \boldsymbol{W}(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}}) \nabla_\beta \boldsymbol{p}(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}}) \\ &\equiv N^{-1} \sum_{i=1}^{N} \nabla_\beta \hat{\boldsymbol{m}}_i' \hat{\boldsymbol{V}}_i^{-1} \hat{\boldsymbol{u}}_i \hat{\boldsymbol{u}}_i' \boldsymbol{V}_i^{-1} \nabla_\beta \hat{\boldsymbol{m}}_i. \end{aligned} \tag{3.10}$$

A fully robust asymptotic variance estimator of $\hat{\boldsymbol{\beta}}$ is

$$\hat{\boldsymbol{A}}^{-1}\hat{\boldsymbol{B}}\hat{\boldsymbol{A}}^{-1}/N. \tag{3.11}$$

This can be used to compute robust standard errors and asymptotic t-statistics.

One important feature of Eq. (3.11) is that it is robust to arbitrary (conditional) serial correlation in $\{u_{it}: t = 1, 2, \dots, T\}$. By definition, $\sum_{t=1}^{T} u_{it} \equiv 0$, so that the u_{it} might generally be expected to exhibit negative serial correlation. In fact, from McCullagh and Nelder, 1989, p. 165), it follows that, under Eqs. (2.1) and (2.2), the correlation between u_{it} and u_{ir}, conditional on $(n_i, \boldsymbol{x}_i)$, is

$$-p_{it}(\boldsymbol{\beta}_0)p_{ir}(\boldsymbol{\beta}_0)/[p_{it}(\boldsymbol{\beta}_0)\{1 - p_{it}(\boldsymbol{\beta}_0)\}p_{ir}(\boldsymbol{\beta}_0)\{1 - p_{ir}(\boldsymbol{\beta}_0)\}]^{1/2}.$$

This particular negative correlation, which is used implicitly in the estimator $\hat{\boldsymbol{A}}^{-1}/N$ of the asymptotic variance of $\hat{\boldsymbol{\beta}}$, need not hold under Eq. (3.1). Consequently, the robust covariance matrix estimator should always be computed; it can produce standard errors smaller or larger than those obtained from $\hat{\boldsymbol{A}}^{-1}/N$.

3.2. Generalized method of moments estimation

In showing that the score of the QCMLE objective function has zero mean conditional on $\boldsymbol{x}_i$ under Eq. (3.1), the key step was to show that $\boldsymbol{u}_i(\boldsymbol{\beta}_0)$ has zero mean conditional on $\boldsymbol{x}_i$. This implies innumerable orthogonality conditions. Let $\boldsymbol{D}(\boldsymbol{x}_i)$ be any $T \times L$ function of $\boldsymbol{x}_i$. Then, assuming the existence of appropriate moments,

$$E[\boldsymbol{D}(\boldsymbol{x}_i)'\boldsymbol{u}_i(\boldsymbol{\beta}_0)] = \boldsymbol{0}; \tag{3.12}$$

this immediately suggests a generalized method of moments (GMM) approach. Actually, in order to obtain estimators that are asymptotically no less efficient than the QCMLE, one must allow $\boldsymbol{D}$ to depend on $\boldsymbol{\beta}$: $\boldsymbol{D}(\boldsymbol{x}_i, \boldsymbol{\beta})$. If $\tilde{\boldsymbol{\beta}}$ is a preliminary consistent estimator of $\boldsymbol{\beta}_0$ – such as the QCMLE – then a minimum chi-square estimator of $\boldsymbol{\beta}_0$ minimizes

$$\left(N^{-1}\sum_{i=1}^{N} \tilde{\boldsymbol{D}}_i'\boldsymbol{u}_i(\boldsymbol{\beta})\right)'\left(N^{-1}\sum_{i=1}^{N} \tilde{\boldsymbol{D}}_i'\tilde{\boldsymbol{u}}_i\tilde{\boldsymbol{u}}_i'\tilde{\boldsymbol{D}}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N} \tilde{\boldsymbol{D}}_i'\boldsymbol{u}_i(\boldsymbol{\beta})\right),$$

where '$\sim$' denotes evaluation at the initial estimate $\tilde{\boldsymbol{\beta}}$. This is the efficient estimator given the choice of the matrix $\boldsymbol{D}(\boldsymbol{x}_i, \boldsymbol{\beta})$.

Since the mechanics of GMM estimation are by now well known, I will not go through the details here. The reader is referred to Newey and McFadden (1994) for a general treatment and Wooldridge (1990b) for the specifics of the current application.

It is important to choose $\boldsymbol{D}(\boldsymbol{x}_i, \boldsymbol{\beta})$ so that the GMM estimator is no less efficient than the QCMLE. This is ensured by including $\boldsymbol{W}(\boldsymbol{x}_i, \boldsymbol{\beta})\nabla_{\beta}\boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta})$ as part of $\boldsymbol{D}(\boldsymbol{x}_i, \boldsymbol{\beta})$, so that the first-order condition for the QCMLE is among the orthogonality conditions for the GMM estimator. The difficulty lies in choosing additional elements of $\boldsymbol{D}(\boldsymbol{x}_i, \boldsymbol{\beta})$ that will produce nontrivial efficiency gains. One possibility is the partitioned matrix

$$\boldsymbol{D}(\boldsymbol{x}_i, \boldsymbol{\beta}) = [\boldsymbol{W}(\boldsymbol{x}_i, \boldsymbol{\beta})\nabla_{\beta}\boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta}) | \nabla_{\beta}\boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta})],$$

so that $\boldsymbol{D}(\boldsymbol{x}_i, \boldsymbol{\beta})$ is a $T \times 2P$ matrix. This adds the orthogonality condition $\mathrm{E}[\nabla_{\beta}\boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)'\boldsymbol{u}_i(\boldsymbol{\beta}_0)] = \boldsymbol{0}$ to the QCMLE first-order condition (giving P overidentifying restrictions), but there is no reason to think this choice is somehow optimal.

An ad hoc approach to choosing additional orthogonality conditions can be avoided by finding the optimal set of instruments. Chamberlain (1992a) studies this problem for a class of unobserved effects models that contains Eq. (3.1) as a special case. Chamberlain's moment conditions are of the form

$$\mathrm{E}[\boldsymbol{Q}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)'\boldsymbol{y}_i|\boldsymbol{x}_i] = \boldsymbol{0}, \tag{3.13}$$

where $\boldsymbol{Q}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)$ is a $T \times L$ matrix such that

$$\boldsymbol{Q}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)'\boldsymbol{\mu}(\boldsymbol{x}_i, \boldsymbol{\beta}_0) \equiv \boldsymbol{0}.$$

Chamberlain characterizes the optimal choice of $\boldsymbol{Q}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)$. Choosing $\boldsymbol{Q}(\boldsymbol{x}_i, \boldsymbol{\beta}) \equiv \boldsymbol{W}(\boldsymbol{x}_i, \boldsymbol{\beta})\nabla_\beta \boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta})$ puts the multinomial QCMLE first-order conditions in the form of Eq. (3.13).

3.3. Specification testing

Testing the conditional mean specification (3.1) can be carried out by testing

$$\mathrm{H}_0\colon \mathrm{E}[\boldsymbol{u}_i(\boldsymbol{\beta}_0)|\boldsymbol{x}_i] = \boldsymbol{0}. \tag{3.14}$$

This is straightforward in a minimum chi-square setting with overidentifying restrictions: one simply uses Hansen's (1982) overidentification test statistic (see also Davidson and MacKinnon, 1993, Section 17.6; and Newey and McFadden, 1994, Section 9.5).

Since the QCMLE is easy to obtain, one may want to obtain relatively simple tests of Eq. (3.1) after QCMLE estimation. Such tests can be obtained along the lines of Newey (1985) and Wooldridge (1990a). If $\boldsymbol{\Lambda}(\boldsymbol{x}_i, \boldsymbol{\beta})$ is a $T \times Q$ matrix depending on $\boldsymbol{x}_i$ and $\boldsymbol{\beta}$, a general class of tests is based on the sample moment

$$N^{-1}\sum_{i=1}^{N} \boldsymbol{\Lambda}(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}})'\boldsymbol{u}_i(\hat{\boldsymbol{\beta}}), \tag{3.15}$$

which should not be statistically different from zero if Eq. (3.1) is true. (The function $\boldsymbol{\Lambda}$ can also depend on other estimated nuisance parameters, but we do not show that here; see Wooldridge (1990b) for details.)

Under H_0, the following expansions are easily seen to hold (for similar reasoning, see Wooldridge, 1990a):

$$\begin{aligned} N^{-1/2}\sum_{i=1}^{N} \boldsymbol{\Lambda}(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}})'\boldsymbol{u}_i(\hat{\beta}) &= N^{-1/2}\sum_{i=1}^{N} \boldsymbol{\Lambda}_i(\boldsymbol{\beta}_0)'\boldsymbol{u}_i(\boldsymbol{\beta}_0) \\ &\quad - \mathrm{E}[\boldsymbol{\Lambda}_i(\boldsymbol{\beta}_0)'\nabla_\beta \boldsymbol{m}_i(\boldsymbol{\beta}_0)]\sqrt{N}(\hat{\beta} - \boldsymbol{\beta}_0) + \mathrm{o_p}(1) \\ &= N^{-1/2}\sum_{i=1}^{N} \{\boldsymbol{\Lambda}_i(\boldsymbol{\beta}_0) - \boldsymbol{W}(\boldsymbol{x}_i,\boldsymbol{\beta}_0)\nabla_\beta \boldsymbol{p}(\boldsymbol{x}_i,\boldsymbol{\beta}_0)\boldsymbol{A}_0^{-1}\boldsymbol{K}_0'\}' \\ &\quad \times \boldsymbol{u}_i(\boldsymbol{\beta}_0) + \mathrm{o_p}(1), \end{aligned}$$

where $\boldsymbol{A}_0 = \mathrm{E}[n_i \nabla_\beta \boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)' \boldsymbol{W}(\boldsymbol{x}_i, \boldsymbol{\beta}_0) \nabla_\beta \boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)]$, $\boldsymbol{\Lambda}_i(\boldsymbol{\beta}_0) \equiv \boldsymbol{\Lambda}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)$, and $\boldsymbol{K}_0 \equiv \mathrm{E}[\boldsymbol{\Lambda}_i(\boldsymbol{\beta}_0)' \nabla_\beta \boldsymbol{m}_i(\boldsymbol{\beta}_0)]$. Thus, let

$$\hat{\boldsymbol{K}} \equiv N^{-1} \sum_{i=1}^{N} \boldsymbol{\Lambda}_i(\hat{\boldsymbol{\beta}})' \nabla_\beta \boldsymbol{m}_i(\hat{\boldsymbol{\beta}}) \tag{3.16}$$

and

$$\hat{\boldsymbol{r}}_i \equiv \hat{\boldsymbol{u}}_i'(\hat{\boldsymbol{\Lambda}}_i - \hat{\boldsymbol{W}}_i \nabla_\beta \hat{\boldsymbol{p}}_i \hat{\boldsymbol{A}}^{-1} \hat{\boldsymbol{K}}'), \quad i = 1, \ldots, N \tag{3.17}$$

(note that $\hat{\boldsymbol{r}}_i$ is a $1 \times Q$ vector). A valid test statistic is obtained as $N -$ SSR from the auxiliary regression

$$1 \text{ on } \hat{\boldsymbol{r}}_i, \quad i = 1, \ldots, N, \tag{3.18}$$

where SSR is just the usual sum of squared residuals. Under the null hypothesis (3.14), $N - \mathrm{SSR} \overset{a}{\sim} \chi^2_Q$, provided there are no redundant columns in $\boldsymbol{\Lambda}_i(\boldsymbol{\beta}_0) - \boldsymbol{W}(\boldsymbol{x}_i, \boldsymbol{\beta}_0) \nabla_\beta \boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta}_0) \boldsymbol{A}_0^{-1} \boldsymbol{K}_0'$.

For example, if we want to test Eq. (3.14) using the extra orthogonality conditions $\mathrm{E}[\nabla_\beta \boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)' \boldsymbol{u}_i(\boldsymbol{\beta}_0)] = \boldsymbol{0}$, then $Q = P$ and $\hat{\boldsymbol{\Lambda}}_i \equiv \nabla_\beta \boldsymbol{p}_i(\hat{\boldsymbol{\beta}})$ in Eqs. (3.16) and (3.17).

4. Conditional linear predictors

This section defines the notion of a conditional linear predictor (or conditional linear projection, CLP) and derives some simple properties; these are convenient for analyses in Sections 5–7. Conditional linear predictors were introduced by Hansen and Richard (1987) in the context of asset pricing relationships; see also Hansen et al. (1988). Here, we use CLPs to obtain orthogonality conditions in the context of certain unobserved effects models.

Unsurprisingly, much of the intuition about linear predictors in an unconditional setting generally carries over to the conditional case. Let $\boldsymbol{y}$ be $J \times 1$, $\boldsymbol{z}$ be $K \times 1$, and $\boldsymbol{w}$ be $I \times 1$. In what follows, $\boldsymbol{z}$ may or may not contain unity as one of its elements. This distinction turns out to be important in the applications. In Section 5, it is inconsequential whether or not unity is included in $\boldsymbol{z}$; in the example of Section 6, unity must be excluded from $\boldsymbol{z}$.

Throughout this section, without stating it explicitly, an expectation is assumed to exist whenever it is written down. Define the following conditional moments:

$$\boldsymbol{\Sigma}_{yz}(\boldsymbol{w}) = \mathrm{E}(\boldsymbol{y}\boldsymbol{z}'|\boldsymbol{w}), \qquad \boldsymbol{\Sigma}_{zz}(\boldsymbol{w}) \equiv \mathrm{E}(\boldsymbol{z}\boldsymbol{z}'|\boldsymbol{w}). \tag{4.1}$$

Assume that $\boldsymbol{\Sigma}_{zz}(\boldsymbol{w})$ is positive definite with probability one (w.p.1.). The following definition holds only w.p.1., but this is left implicit throughout.

Definition 4.1. Let $\boldsymbol{y}$, $\boldsymbol{z}$, and $\boldsymbol{w}$ be defined as above. The *linear predictor of* $\boldsymbol{y}$ *on* $\boldsymbol{z}$, *conditional on* $\boldsymbol{w}$, is defined to be

$$L(\boldsymbol{y}|\boldsymbol{z};\ \boldsymbol{w}) \equiv \boldsymbol{\Sigma}_{yz}(\boldsymbol{w})\boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{w})\boldsymbol{z} \equiv \boldsymbol{C}_0(\boldsymbol{w})\boldsymbol{z}, \tag{4.2}$$

where $\boldsymbol{C}_0(\boldsymbol{w})$ is the $J \times K$ matrix $\boldsymbol{C}_0(\boldsymbol{w}) \equiv \boldsymbol{\Sigma}_{yz}(\boldsymbol{w})\boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{w})$. □

Note that $L(\boldsymbol{y}|\boldsymbol{z};\ \boldsymbol{w})$ is always linear in $\boldsymbol{z}$, but is generally a nonlinear function of $\boldsymbol{w}$. When the context is clear, $L(\boldsymbol{y}|\boldsymbol{z};\ \boldsymbol{w})$ is simply called a conditional linear predictor. The difference between $\boldsymbol{y}$ and its CLP has zero orthogonality properties that are immediate extensions from unconditional linear predictor theory.

Lemma 4.1. Let $\boldsymbol{y}$, $\boldsymbol{z}$, *and* $\boldsymbol{w}$ *be as in Definition 4.1. Define*

$$\boldsymbol{u} \equiv \boldsymbol{y} - L(\boldsymbol{y}|\boldsymbol{z};\ \boldsymbol{w}) = \boldsymbol{y} - \boldsymbol{C}_0(\boldsymbol{w})\boldsymbol{z}. \tag{4.3}$$

Then

$$\mathrm{E}(\boldsymbol{u}\boldsymbol{z}'|\boldsymbol{w}) = \boldsymbol{0}. \tag{4.4}$$

Proof. $\boldsymbol{u}\boldsymbol{z}' = [\boldsymbol{y} - \boldsymbol{C}_0(\boldsymbol{w})\boldsymbol{z}]\boldsymbol{z}' = \boldsymbol{y}\boldsymbol{z}' - \boldsymbol{C}_0(\boldsymbol{w})\boldsymbol{z}\boldsymbol{z}'$, so that

$$\mathrm{E}(\boldsymbol{u}\boldsymbol{z}'|\boldsymbol{w}) = \mathrm{E}(\boldsymbol{y}\boldsymbol{z}'|\boldsymbol{w}) - \boldsymbol{C}_0(\boldsymbol{w})\mathrm{E}(\boldsymbol{z}\boldsymbol{z}'|\boldsymbol{w}) = \boldsymbol{\Sigma}_{yz}(\boldsymbol{w}) - \boldsymbol{\Sigma}_{yz}(\boldsymbol{w})\boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{w})\boldsymbol{\Sigma}_{zz}(\boldsymbol{w}) = \boldsymbol{0}. \quad \square$$

The next corollary formalizes the class of orthogonality conditions that can be used in estimating the parameters of a conditional linear predictor. It is a simple application of the law of iterated expectations.

Corollary 4.1. Let $\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w}$, *and* $\boldsymbol{u}$ *be as in Lemma 4.1, and let* $\boldsymbol{D}(\boldsymbol{w})$ *be a* $JK \times L$ *random matrix. Then*

$$\mathrm{E}[\boldsymbol{D}(\boldsymbol{w})'(\boldsymbol{z} \otimes \boldsymbol{I}_J)\boldsymbol{u}] = \mathrm{E}[\boldsymbol{D}(\boldsymbol{w})'\ \mathrm{vec}\{\boldsymbol{u}\boldsymbol{z}'\}] = \boldsymbol{0}. \tag{4.5}$$

Suppose now that $\boldsymbol{C}_0(\boldsymbol{w}) \equiv \boldsymbol{C}(\boldsymbol{w}, \boldsymbol{\theta}_0)$, where $\boldsymbol{C}(\boldsymbol{w}, \boldsymbol{\theta})$ is a known function of $\boldsymbol{w}$ and the $P \times 1$ parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Then, for a matrix function $\boldsymbol{D}(\boldsymbol{w})$ as defined in Corollary 4.1, $\boldsymbol{\theta}_0$ solves (perhaps not uniquely) the system of equations

$$\mathrm{E}[\boldsymbol{D}(\boldsymbol{w})'(\boldsymbol{z} \otimes \boldsymbol{I}_J)\{\boldsymbol{y} - \boldsymbol{C}(\boldsymbol{w}, \boldsymbol{\theta})\}] \equiv \boldsymbol{0}. \tag{4.6}$$

Eq. (4.6) can be exploited to obtain a variety of consistent estimators of $\boldsymbol{\theta}_0$. In the applications in Sections 5–7, not all of the vector $\boldsymbol{w}$ is observed (in particular, $\boldsymbol{w}$ contains unobserved effects as well as observed conditioning variables). The CLP is useful for estimating $\boldsymbol{\theta}_0$ only if $\boldsymbol{C}(\boldsymbol{w}, \boldsymbol{\theta}_0)$ does not depend on the unobserved elements of $\boldsymbol{w}$, as in the subsequent examples.

5. A count-type model under mean and variance assumptions

The methods in Section 3 require only that the conditional mean in the latent variable model be correctly specified. These estimators have satisfying robustness properties, but may have large asymptotic variances. As mentioned earlier, Chamberlain (1992a) derives the variance lower bound of estimators that use only Eq. (3.1). But achieving this lower bound via optimal instrumental variables estimation generally requires nonparametric estimation of a very high dimensional conditional expectation.

A different way to improve efficiency is to make conditional variance and covariance assumptions in the unobserved effects model and use these assumptions in estimating $\boldsymbol{\beta}_0$. If y_{it} is a count variable, we might wish to impose a constant variance–mean ratio, as is popular in the count data literature. Nonlinear unobserved effects models that use full distributional assumptions assume conditional independence in the y_{it} across time, something we might want to exploit in estimation without making full distributional assumptions.

The following model is motivated by, but certainly not restricted to, count data. For $t = 1, \ldots, T$, assume that

$$\mathrm{E}(y_{it}|\boldsymbol{x}_i, \phi_i, \psi_i) = \phi_i \mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0), \tag{5.1}$$

$$\mathrm{Var}(y_{it}|\boldsymbol{x}_i, \phi_i, \psi_i) = \psi_i \mathrm{E}(y_{it}|\boldsymbol{x}_i, \phi_i, \psi_i) = \psi_i \phi_i \mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0), \tag{5.2}$$

$$\mathrm{Cov}(y_{it}, y_{ir}|\boldsymbol{x}_i, \phi_i, \psi_i) = 0, \quad t \neq r. \tag{5.3}$$

This model has two unobserved effects, one that multiplies the mean and the other that determines the variance-to-mean ratio:

$$\mathrm{Var}(y_{it}|\boldsymbol{x}_i, \phi_i, \psi_i)/\mathrm{E}(y_{it}|\boldsymbol{x}_i, \phi_i, \psi_i) = \psi_i. \tag{5.4}$$

Models with a constant variance-to-mean ratio are known as *log-linear models* in the generalized linear models literature (e.g. McCullagh and Nelder, 1989, Chapter 6). Assumption (5.2) relaxes this assumption by allowing the variance-to-mean ratio to be different for each i. This means that some cross-section units can display underdispersion and others overdispersion.

Assumptions (5.1)–(5.3) are notably more flexible than the first two moments of all of the fixed effects models estimated by HHG. The fixed effects Poisson model imposes $\psi_i \equiv 1$. The fixed effects negative binomial model of HHG, mentioned in Section 2, imposes $\psi_i = 1 + \phi_i$, which rules out underdispersion for all individuals and ties the amount of overdispersion directly to the mean effect. HHG introduced two unobserved effects in their negative binomial model, μ_i and ϕ_i, but it is easily seen that only the ratio $1/\delta_i \equiv \exp(\mu_i)/\phi_i$ actually appears in their model. In addition, Eq. (5.3) is weaker than the independence assumption imposed by HHG, and no distributional assumption is made.

The variance assumption (5.2) is also more general than that implied by a common parameterization of the gamma distribution (for example, Rao, 1973, p. 164); in this case, $\psi_i = \phi_i$. Therefore, this model can be usefully applied to continuous, nonnegative responses.

The orthogonality condition

$$\mathrm{E}[\boldsymbol{u}_i(\boldsymbol{\beta}_0)|\boldsymbol{x}_i] = \boldsymbol{0}, \tag{5.5}$$

where $\boldsymbol{u}_i(\boldsymbol{\beta}) \equiv \boldsymbol{y}_i - \boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta})n_i$, was derived under Eq. (3.1) only. Therefore, the additional assumptions (5.2) and (5.3) are useful only if they imply orthogonality conditions in addition to Eq. (5.5). In fact, assumptions (5.1)–(5.3) also imply that

$$\mathrm{E}[n_i\boldsymbol{u}_i(\boldsymbol{\beta}_0)|\boldsymbol{x}_i] = \boldsymbol{0}, \tag{5.6}$$

where $n_i \equiv \sum_{t=1}^{T} y_{it}$, as before. Condition (5.6) is a simple consequence of the form of the linear predictor of y_{it} on $(1, n_i)'$, conditional on $(\boldsymbol{x}_i, \phi_i, \psi_i)$. From Definition 4.1, for each t

$$L(y_{it}|1,\, n_i;\, \boldsymbol{x}_i, \phi_i, \psi_i) = \mathrm{E}(y_{it}|\boldsymbol{x}_i, \phi_i, \psi_i) + \frac{\mathrm{Cov}(y_{it}, n_i|\boldsymbol{x}_i, \phi_i, \psi_1)}{\mathrm{Var}(n_i|\boldsymbol{x}_i, \phi_i, \psi_i)} \times (n_i - \mathrm{E}(n_i|\boldsymbol{x}_i, \phi_i, \psi_i)) \tag{5.7}$$

$$= \mathrm{E}(y_{it}|\boldsymbol{x}_i, \phi_i, \psi_i) + \frac{\mathrm{Var}(y_{it}|\boldsymbol{x}_i, \phi_i, \psi_i)}{\mathrm{Var}(n_i|\boldsymbol{x}_i, \phi_i, \psi_i)}(n_i - \mathrm{E}(n_i|\boldsymbol{x}_i, \phi_i, \psi_i))$$

$$= \phi_i\mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0)\frac{\psi_i\phi_i\mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0)}{\sum_{r=1}^{T}\psi_i\phi_i\mu(\boldsymbol{x}_{ir}, \boldsymbol{\beta}_0)}\left(n_i - \sum_{r=1}^{T}\phi_i\mu(\boldsymbol{x}_{ir}, \boldsymbol{\beta}_0)\right)$$

$$= \frac{\mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0)}{\sum_{r=1}^{T}\mu(\boldsymbol{x}_{ir}, \boldsymbol{\beta}_0)}n_i, \tag{5.8}$$

which does not depend on ϕ_i or ψ_i. It follows from Lemma 4.1 that $\mathrm{E}[n_i\{\boldsymbol{y}_i - \boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)n_i\}|\boldsymbol{x}_i, \phi_i, \psi_i] = 0$, which implies Eq. (5.6). It is important to see that, as in Section 3, Eq. (5.8) does *not* generally equal $\mathrm{E}(y_{it}|n_i, \boldsymbol{x}_i, \phi_i, \psi_i)$.

Eqs. (5.5) and (5.6) imply orthogonality conditions of the form

$$\mathrm{E}[\boldsymbol{D}(\boldsymbol{x}_i)'\{(1, n_i)' \otimes \boldsymbol{I}_T\}\boldsymbol{u}_i(\boldsymbol{\beta}_0)] = \boldsymbol{0}, \tag{5.9}$$

where $\boldsymbol{D}(\boldsymbol{x}_i)$ is any $2T \times L$ matrix function of $\boldsymbol{x}_i$. These orthogonality conditions can be used in GMM estimation.

Interestingly, the multivariate nonlinear least-squares (MNLS) estimator $\hat{\boldsymbol{\beta}}$, which solves

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \{\boldsymbol{y}_i - \boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta})n_i\}'\{\boldsymbol{y}_i - \boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta})n_i\}/2, \tag{5.10}$$

has a first-order condition of the form (5.9): $\mathrm{E}[n_i \nabla_\beta \boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)'\boldsymbol{u}_i(\boldsymbol{\beta}_0)] = \boldsymbol{0}$. Therefore, under Eqs. (5.1)–(5.3) and standard regularity conditions, $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}_0$ and asymptotically normally distributed. This estimator is *not* generally consistent for $\boldsymbol{\beta}_0$ under the conditional mean assumption (5.1) only. This is because Eq. (5.1) alone does not imply Eq. (5.6), which is the basis for the MNLS consistency result. The asymptotic variance of $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ takes the form $\boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1}$, with consistent estimators of $\boldsymbol{A}_0$ and $\boldsymbol{B}_0$ given by

$$\hat{\boldsymbol{A}} \equiv N^{-1} \sum_{i=1}^{N} \nabla_\beta \hat{\boldsymbol{m}}_i' \nabla_\beta \hat{\boldsymbol{m}}_i \tag{5.11}$$

and

$$\hat{\boldsymbol{B}} \equiv N^{-1} \sum_{i=1}^{N} \nabla_\beta \hat{\boldsymbol{m}}_i' \hat{\boldsymbol{u}}_i \hat{\boldsymbol{u}}_i' \nabla_\beta \hat{\boldsymbol{m}}_i, \tag{5.12}$$

where $\nabla_\beta \hat{\boldsymbol{m}}_i \equiv \nabla_\beta \boldsymbol{m}_i(\hat{\boldsymbol{\beta}}) \equiv \nabla_\beta \boldsymbol{p}_i(\hat{\boldsymbol{\beta}})n_i$. This is easily extended to weighted least squares with $T \times T$ weighting matrix $\hat{\boldsymbol{G}}_i(\boldsymbol{x}_i) \equiv \boldsymbol{G}(\boldsymbol{x}_i, \hat{\gamma})$, which can depend on $\boldsymbol{x}_i$ (and some estimated parameters $\hat{\gamma}$) but not n_i; see Wooldridge (1990b) for details.

Because Eq. (5.6) does not generally hold under Eq. (5.1), these additional orthogonality conditions can be used to test the variance and covariance assumptions (5.2) and (5.3). One simple way to do this is to compare the multinomial QCMLE and the MNLS estimator via a Hausman (1978) test. However, neither estimator is relatively efficient under Eqs. (5.1)–(5.3), so a robust form of the test is needed. A regression-based approach, which requires estimation of the QCMLE only, is obtained from Wooldridge (1991). Let $\hat{\boldsymbol{u}}_i, \nabla_\beta \hat{\boldsymbol{m}}_i$, and $\hat{\boldsymbol{V}}_i$ be defined as in Section 3, evaluated at the QCMLE, $\hat{\boldsymbol{\beta}}$.

Define the weighted quantities $\tilde{\boldsymbol{u}}_i \equiv \hat{\boldsymbol{V}}_i^{-1/2}\hat{\boldsymbol{u}}_i, \nabla_\beta\tilde{\boldsymbol{m}}_i \equiv \hat{\boldsymbol{V}}_i^{-1/2}\nabla_\beta\hat{\boldsymbol{m}}_i$, and $\tilde{\Lambda}_i \equiv \hat{\boldsymbol{V}}_i^{1/2}\nabla_\beta\hat{\boldsymbol{m}}_i$. The robust Hausman test is easily computed by first orthogonalizing $\tilde{\Lambda}_i$ with respect to $\nabla_\beta\tilde{\boldsymbol{m}}_i$. Let $\tilde{\boldsymbol{E}}_i$ be the $T \times P$ matrix residuals from the matrix regression

$$\tilde{\Lambda}_i \text{ on } \nabla_\beta\tilde{\boldsymbol{m}}_i, \quad i = 1, \ldots, N. \tag{5.13}$$

Then compute $H \equiv N -$ SSR from the regression

$$1 \text{ on } \tilde{\boldsymbol{u}}_i'\tilde{\boldsymbol{E}}_i, \quad i = 1, \ldots, N; \tag{5.14}$$

under Eqs. (5.5) and (5.6), $H \overset{a}{\sim} \chi^2_P$. If the specification tests in Section 3 fail to reject Eq. (3.1), but Eq. (5.14) rejects Eqs. (5.1)–(5.3), this can be taken as evidence that at least one of the second moment assumptions (5.2) and (5.3) is violated.

If a minimum chi-square approach is used, specification tests are available from the overidentification test statistics. A useful GMM estimator stacks the orthogonality conditions of QCMLE and MNLS, and then the overidentification test can be used to test assumptions (5.1)–(5.3) (see Wooldridge, 1990b for details).

It is important to remember that the moment assumptions (5.1)–(5.3) encompass HHG's fixed effects negative binomial model. Consequently, a rejection of either Eqs. (5.5) and (5.6) based on H or any other test statistic necessarily implies misspecification of the FENB specification. A rejection implies some failure of Eqs. (5.1)–(5.3), so one needs to work harder in specifying $\mathrm{E}(\boldsymbol{y}_i|\boldsymbol{x}_i, \phi_i, \psi_i)$ or $\mathrm{Var}(\boldsymbol{y}_i|\boldsymbol{x}_i, \phi_i, \psi_i)$.

6. A gamma-type model under mean and variance assumptions

The variance assumption (5.2) is known to fail for some popular continuous distributions. For example, the parameterization of the gamma distribution used in the quasi-likelihood literature (for example, GMT, p. 685) implies that the conditional variance of y_{it} is proportional to the *square* of the conditional mean. This is also true for a natural parameterization of the lognormal distribution. This motivates the following model for all t:

$$\mathrm{E}(y_{it}|\boldsymbol{x}_i, \phi_i) = \phi_i\mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0), \tag{6.1}$$

$$\mathrm{Var}(y_{it}|\boldsymbol{x}_i, \phi_i) = \sigma_0^2[\mathrm{E}(y_{it}|\boldsymbol{x}_i, \phi_i)]^2 = \sigma_0^2[\phi_i\mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0)]^2, \tag{6.2}$$

$$\mathrm{Cov}(y_{it}, y_{ir}|\boldsymbol{x}_i, \phi_i) = 0, \quad t \neq r. \tag{6.3}$$

Assumption (6.1) is essentially the same as Eq. (5.1), while Eq. (6.3) is essentially Eq. (5.3). Eq. (6.2) corresponds to what is sometimes referred to as a *constant coefficient of variation model* (e.g. McCullagh and Nelder, 1989, Chapter 8) because the ratio of the standard deviation to the mean is constant. Ideally, the parameter σ_0^2 could be replaced by an unobserved effect that depends on i, but I do not know how to do that using the conditional linear predictor approach.

As before, under Eq. (6.1) $\boldsymbol{\beta}_0$ can be estimated using the methods in Section 3. However, assumptions (6.2) and (6.3) expand the list of orthogonality conditions. As far as I know, there has been no work analyzing models such as Eqs. (6.1)–(6.3) in an unobserved components, panel data setting, with or without distributional assumptions. This is probably because, when y_{it} is a non-negative continuous random variable, most researchers use $\log(y_{it})$ in a linear fixed effects model.

Define $\boldsymbol{p}(\boldsymbol{x}_i, \boldsymbol{\beta})$,$n_i$, and $\boldsymbol{u}_i(\boldsymbol{\beta})$ as in the previous sections. Additional orthogonality conditions are obtained by computing the linear predictor of y_{it} on n_i, conditional on $(\boldsymbol{x}_i, \phi_i)$. For each $t = 1, \dots, T$,

$$L(y_{it}|n_i; \boldsymbol{x}_i, \phi_i) \equiv \frac{\mathrm{E}(y_{it}n_i|\boldsymbol{x}_i, \phi_i)}{\mathrm{E}(n_i^2|\boldsymbol{x}_i, \phi_i)} n_i. \tag{6.4}$$

To show that Eq. (6.4) does not depend on ϕ_i, note that

$$\begin{aligned}
\mathrm{E}(y_{it}n_i|\boldsymbol{x}_i, \phi_i) &= \mathrm{Cov}(y_{it}, n_i|\boldsymbol{x}_i, \phi_i) + \mathrm{E}(y_{it}|\boldsymbol{x}_i, \phi_i)\mathrm{E}(n_i|\boldsymbol{x}_i, \phi_i) \\
&= \mathrm{Var}(y_{it}|\boldsymbol{x}_i, \phi_i) + [\phi_i\mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0)]\left(\sum_{r=1}^{T} \phi_i\mu(\boldsymbol{x}_{ir}, \boldsymbol{\beta}_0)\right) \\
&= \phi_i^2\left(\sigma_0^2[\mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0)]^2 + \mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0)\left(\sum_{r=1}^{T} \mu(\boldsymbol{x}_{ir}, \boldsymbol{\beta}_0)\right)\right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathrm{E}(n_i^2|\boldsymbol{x}_i, \phi_i) &= \mathrm{Var}(n_i|\boldsymbol{x}_i, \phi_i) + [\mathrm{E}(n_i|\boldsymbol{x}_i, \phi_i)]^2 \\
&= \phi_i^2\left(\sigma_0^2\sum_{r=1}^{T} [\mu(\boldsymbol{x}_{ir}, \boldsymbol{\beta}_0)]^2 + \left(\sum_{r=1}^{T} \mu(\boldsymbol{x}_{ir}, \boldsymbol{\beta}_0)\right)^2\right).
\end{aligned}$$

Therefore,

$$L(y_{it}|n_i; \boldsymbol{x}_i, \phi_i) \equiv q_t(\boldsymbol{x}_i, \boldsymbol{\theta}_0)n_i, \tag{6.5}$$

where $\boldsymbol{\theta}_0 \equiv (\boldsymbol{\beta}_0', \sigma_0^2)'$ and

$$q_t(\boldsymbol{x}_i,\boldsymbol{\theta}_0) \equiv \frac{\sigma_0^2[\mu(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0)]^2 + \mu(\boldsymbol{x}_{it},\boldsymbol{\beta}_0)\ (\sum_{r=1}^{T}\mu(\boldsymbol{x}_{ir}, \boldsymbol{\beta}_0))}{\sigma_0^2\sum_{r=1}^{T}\ [\mu(\boldsymbol{x}_{ir}, \boldsymbol{\beta}_0)]^2 + (\sum_{r=1}^{T}\mu(\boldsymbol{x}_{ir}, \boldsymbol{\beta}_0))^2}. \tag{6.6}$$

Note that, for all $\boldsymbol{\theta}$, $\sum_{t=1}^{T} q_t(\boldsymbol{x}_i, \boldsymbol{\theta}) = 1$.

In terms of the vector $\boldsymbol{y}_i$, Eq. (6.5) is expressed as

$$L(\boldsymbol{y}_i|n_i; \boldsymbol{x}_i, \phi_i) = \boldsymbol{q}(\boldsymbol{x}_i, \boldsymbol{\theta}_0)n_i, \tag{6.7}$$

where $\boldsymbol{q}(\boldsymbol{x}_i,\boldsymbol{\theta}) \equiv [q_1(\boldsymbol{x}_i, \boldsymbol{\theta}), \dots, q_T(\boldsymbol{x}_i, \boldsymbol{\theta})]'$ is $T \times 1$. Let $\boldsymbol{v}_i(\boldsymbol{\theta}) \equiv \boldsymbol{y}_i - \boldsymbol{q}(\boldsymbol{x}_i,\boldsymbol{\theta})n_i$. Then Eqs. (6.1), (6.2) and (6.3) imply joint orthogonality conditions of the form

$$\mathrm{E}[\boldsymbol{D}_1(\boldsymbol{x}_i)'\boldsymbol{u}_i(\boldsymbol{\beta}_0)] = \boldsymbol{0}, \tag{6.8}$$

$$\mathrm{E}[\boldsymbol{D}_2(\boldsymbol{x}_i)'n_i\boldsymbol{v}_i(\boldsymbol{\theta}_0)] = \boldsymbol{0}, \tag{6.9}$$

where $\boldsymbol{D}_1(\boldsymbol{x}_i)$ is $T \times L_1$ and $\boldsymbol{D}_2(\boldsymbol{x}_i)$ is $T \times L_2$.

For example, weighted MNLS estimators, which solve

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \{\boldsymbol{y}_i - \boldsymbol{q}(\boldsymbol{x}_i, \boldsymbol{\theta})n_i\}'\boldsymbol{G}(\boldsymbol{x}_i, \hat{\gamma})\{\boldsymbol{y}_i - \boldsymbol{q}(\boldsymbol{x}_i, \boldsymbol{\theta})n_i\}, \tag{6.10}$$

are generally consistent and asymptotically normally distributed for $\boldsymbol{\beta}_0$ and σ_0^2. Further, given one such estimator, it is straightforward to stack WMNLS orthogonality conditions along with the multinomial QCMLE orthogonality conditions to obtain a more efficient minimum chi-square estimator. The details are omitted.

7. Models with parametric serial correlation

For some applications, the zero covariance assumptions (5.3) and (6.3) might be too restrictive (but one should remember that these are conditional on latent effects; conditional on $\boldsymbol{x}_i$ only, the y_{it} are correlated). The methods of Section 3 are available without any assumptions on the variances and covariances, but sometimes we might want to model the serial correlation parametrically and use this information in estimating $\boldsymbol{\beta}_0$.

For the model in Section 6, it is straightforward to relax the zero covariance assumption. In fact, Eqs. (6.2) and (6.3) can be replaced with the more general assumption

$$\mathrm{Var}(\boldsymbol{y}_i|\boldsymbol{x}_i, \phi_i) = \phi_i^2\boldsymbol{\Omega}(\boldsymbol{x}_i, \boldsymbol{\delta}_0), \tag{7.1}$$

where $\boldsymbol{\Omega}(\boldsymbol{x}_i,\boldsymbol{\delta})$ is a $T \times T$ positive-definite variance function. The conditional linear prediction argument used in Section 6 still eliminates ϕ_i. In fact, letting $\boldsymbol{\Omega}_t(\boldsymbol{x}_i,\boldsymbol{\delta})$ denote the tth row of $\boldsymbol{\Omega}(\boldsymbol{x}_i,\boldsymbol{\delta})$ and $\boldsymbol{j}_T \equiv (1,1,\dots,1)'$, $L(y_{it}|n_i;\, \boldsymbol{x}_i, \phi_i)$ is given by Eq. (6.6) with

$$q_t(\boldsymbol{x}_i,\boldsymbol{\theta}_0) \equiv \frac{\boldsymbol{\Omega}_t(\boldsymbol{x}_i,\boldsymbol{\delta}_0)\,\boldsymbol{j}_T + \mu(\boldsymbol{x}_{it},\boldsymbol{\beta}_0)(\sum_{r=1}^{T}\mu(\boldsymbol{x}_{ir},\boldsymbol{\beta}_0))}{\boldsymbol{j}_T'\boldsymbol{\Omega}(\boldsymbol{x}_i,\boldsymbol{\delta}_0)\,\boldsymbol{j}_T + (\sum_{r=1}^{T}\mu(\boldsymbol{x}_{ir},\boldsymbol{\beta}_0))^2}. \tag{7.2}$$

The model that combines Eqs. (6.1) and (7.1) allows for serial correlation and a variety of variance functions. For example, a gamma-type model with constant AR(1) serial correlation would take

$$\boldsymbol{\Omega}(\boldsymbol{x}_i,\boldsymbol{\delta}_0) \equiv \sigma_0^2\boldsymbol{\Delta}(\boldsymbol{x}_i,\boldsymbol{\beta}_0)\boldsymbol{R}_T(\rho_0)\boldsymbol{\Delta}(\boldsymbol{x}_i,\boldsymbol{\beta}_0), \tag{7.3}$$

where

$$\boldsymbol{\Delta}(\boldsymbol{x}_i,\boldsymbol{\beta}_0) \equiv \mathrm{diag}\{\mu(\boldsymbol{x}_{i1},\boldsymbol{\beta}_0),\dots,\, \mu(\boldsymbol{x}_{iT},\boldsymbol{\beta}_0)\} \tag{7.4}$$

and $\boldsymbol{R}_T(\rho)$ is the $T \times T$ matrix with (r,t)th element $\rho^{|r-t|}$. This model maintains Eq. (6.2) but relaxes Eq. (6.3) to $\mathrm{Cov}(y_{it}, y_{ir}|\boldsymbol{x}_i, \phi_i) = \phi_i^2\sigma_0^2\rho_0^{|r-t|}\mu(\boldsymbol{x}_{it},\boldsymbol{\beta}_0)\mu(\boldsymbol{x}_{ir},\boldsymbol{\beta}_0)$.

We can also allow for serial correlation in models where the variance is proportional to the mean (as in Section 5), but the individual dispersion is restricted in this case. A model with constant AR(1) serial correlation chooses $\boldsymbol{\Omega}(\boldsymbol{x}_i,\boldsymbol{\delta}_0)$ as in Eq. (7.3), except that

$$\boldsymbol{\Delta}(\boldsymbol{x}_i,\boldsymbol{\beta}_0) \equiv \mathrm{diag}\{[\mu(\boldsymbol{x}_{i1},\boldsymbol{\beta}_0)]^{1/2},\dots,[\mu(\boldsymbol{x}_{iT},\boldsymbol{\beta}_0)]^{1/2}\}. \tag{7.5}$$

In terms of model (5.1)–(5.3), Eq. (5.2) has been maintained and Eq. (5.3) has been relaxed at the cost of imposing $\psi_i \equiv \sigma_0^2\phi_i$.

The identification issue in these more complicated models warrants some attention. For example, in Eqs. (6.1), (7.3) and (7.5), it can be seen that σ_0^2 is not identified if $\rho_0 = 0$. In this case the model of Section 5 applies, and a separate variance effect ψ_i can vary independently of ϕ_i.

8. Concluding remarks

This paper offers some distribution-free estimators for multiplicative panel data models. The estimators of Section 3 require only a conditional mean assumption. Perhaps the most practical result is that the fixed effects Poisson estimator is fully robust in the sense that only the structural conditional mean assumption, given in Eq. (3.1), is needed for consistency and asymptotic normality. The robust variance matrix estimate is easy to obtain, and specification testing is fairly straightforward. These results have already been used in several published and unpublished studies, including Papke (1991), Hausman et al. (1995), Page (1995), Gordy (1996), and Crépon and Duguet (1997). The recent proof by Hahn (1997) that the FEP estimator achieves the semiparametric efficiency bound under Eqs. (2.1) and (2.2) provides further justification for the FEP estimator.

A GMM approach that adds orthogonality conditions to the FEP orthogonality conditions might prove useful in future studies, both for enhancing efficiency and providing convenient specification testing.

Sections 5–7 demonstrate how additional orthogonality conditions can be derived from certain models that specify the conditional second moments in addition to the conditional mean. The model in Section 5 extends the mean and variance assumptions for well-known count distributions, and Section 6 studies a variance assumption suited to nonnegative, continuously distributed response variables. Section 7 shows how parametric models of serial correlation can be used in estimation. In each section, no distributional assumptions are needed for consistent estimation: the estimators are robust to assumptions not explicitly imposed.

Chamberlain's (1992a) approach for deriving efficiency bounds can apparently be applied to models such as Eqs. (5.1)–(5.3) by augmenting his orthogonality conditions with orthogonality conditions for the squares and cross products of the explained variable. We leave this to future research.

Finally, the models in Sections 3, 5–7 assume that $\boldsymbol{x}_{it}$ is strictly exogenous conditional on the latent variable or variables. This can be violated especially when there is feedback from y_{it} to $\boldsymbol{x}_{ir}$, $r > t$ (that is, if $\{y_{it}\}$ Granger-causes $\{\boldsymbol{x}_{it}\}$). While strict exogeneity is natural for certain explanatory variables, it is difficult to justify in general. For example, in HHG's patents-R&D application, the number of patents awarded in one year could affect subsequent R&D expenditures. If $\boldsymbol{x}_{it}$ is not strictly exogenous, all of the estimators offered here are inconsistent. In the case of a conditional mean specification, Chamberlain (1992b) and Wooldridge (1997) suggest orthogonality conditions that can be used in method of moments procedures.

Acknowledgements

This is a revised version of MIT Department of Economics Working Paper No. 564. I am grateful to two anonymous referees, an associate editor, Zvi Griliches, Lars Hansen, Jerry Hausman, Danny Quah, Tom Stoker, and especially Gary Chamberlain and Whitney Newey for very helpful comments, suggestions, and extensions.

References

Andersen, E.B., 1970. Asymptotic properties of conditional maximum likelihood estimators. Journal of the Royal Statistical Society Series B 32, 283–301.

Chamberlain, G., 1980. Analysis of covariance with qualitative data. Review of Economic Studies 47, 225–238.

Chamberlain, G., 1982. Multivariate regression models for panel data. Journal of Econometrics 18, 5–46.

Chamberlain, G., 1984. Panel data. In: Griliches, Z., Intriligator, M. (Eds.), Handbook of Econometrics, vol. I, North-Holland, Amsterdam.

Chamberlain, G., 1992a. Efficiency bounds for semiparametric regression. Econometrica 60, 567–596.

Chamberlain, G., 1992b. Comment: sequential moment restrictions in panel data. Journal of Business and Economic Statistics 10, 20–26.

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

Davidson, R., MacKinnon, J.G., 1993. Estimation and Inference in Econometrics. Oxford University Press, New York.

Gordy, M.B., 1996. Hedging winner's curse with multiple bids: evidence from the portuguese treasury bill auction. Mimeo. Division of Research and Statistics, Federal Reserve Board, Washington, DC.

Gourieroux, C., Monfort, A., Trognon, C., 1984. Pseudo-maximum likelihood Methods: theory. Econometrica 52, 681–700.

Hahn, J., 1997. A note on efficient semiparametric estimation of some exponential panel models. Econometric Theory 13, 583–588.

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

Hansen, L.P., Richard, S.F., 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. Econometrica 55, 587–613.

Hansen, L.P., Heaton, J., Ogaki, M., 1988. Efficiency bounds implied by multiperiod conditional moment restrictions. Journal of the American Statistical Association 83, 663–871.

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46, 1251–1271.

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

Hausman, J.A., Leonard, G., McFadden, D.L., 1995. A utility-consistent, combined discrete choice and count data model. Assessing recreational use losses due to natural resource damage. Journal of Public Economics 56, 1–30.

Johnson, N.L., Kotz, S., 1969. Discrete Distributions. Wiley, New York.

McCullagh P., Nelder, J.A., 1989. Generalized Linear Models. Chapman & Hall, New York.

Newey, W.K., 1985. Maximum likelihood specification testing and conditional moment tests. Econometrica 53, 1047–1070.

Newey, W.K., McFadden, D.L., 1994. Large sample estimation and hypothesis testing. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, Amsterdam, pp. 2111–2245.

Page, M., 1995. Racial and ethnic discrimination in urban housing markets: evidence from a recent audit study. Journal of Urban Economics 38, 183–206.

Palmgren, J., 1981. The Fisher information matrix for log-linear models arguing conditionally in the observed explanatory variables. Biometrika 68, 563–566.

Papke, L.E., 1991. Interstate business tax differentials and new firm location. Journal of Public Economics 45, 47–68.

Rao, C.R., 1973. Linear Statistical Inference and its Applications, 2nd Ed. Wiley, New York.

Wooldridge, J.M., 1990a. A unified approach to robust, regression-based specification tests. Econometric Theory 6, 17–43.

Wooldridge, J.M., 1990b. Distibution-free estimation of some nonlinear panel data models. MIT Department of Economics working paper No. 564.

Wooldridge, J.M., 1991. Specification testing and quasi-maximum likelihood estimation. Journal of Econometrics 48, 29–55.

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