


MIT LIBRARIES



3 9080 00701592 5





Digitized by the Internet Archive
in 2011 with funding from
Boston Library Consortium Member Libraries

<http://www.archive.org/details/distributionfree00wool>

HB31
.M415
no. 564

**working paper
department
of economics**

NONLINEAR PANEL DATA MODELS

Jeffrey M. Wooldridge

Number 564

September 1990

**massachusetts
institute of
technology**

**50 memorial drive
cambridge, mass. 02139**

DISTRIBUTION-FREE ESTIMATION OF SOME
NONLINEAR PANEL DATA MODELS

Jeffrey M. Wooldridge

Number 564

September 1990

M.I.M. JAMES
DEC 26 1990
FBI

DISTRIBUTION-FREE ESTIMATION OF SOME
NONLINEAR PANEL DATA MODELS

Jeffrey M. Wooldridge
Department of Economics
Massachusetts Institute of Technology, E52-262C
Cambridge, MA 02139
(617) 253-3488
JMWOOLDR@ATHENA.MIT.EDU

July 1990
Latest Revision: September 1990

Abstract

The notion of a conditional linear predictor is used as a distribution-free method for eliminating the individual-specific effects in a class of nonlinear, unobserved components panel data models. The methodology is applied to a general count model, which allows for individual dispersion in addition to an individual mean effect. As a corollary of the general results, the multinomial quasi-conditional maximum likelihood estimator is shown to be consistent and asymptotically normal when only the first two moments in the unobserved effects model have been correctly specified. This has important implications for analyzing count data in panel contexts. Simple, robust specification tests for this class of count models are also developed. A second example covers the case where the variance is proportional to the square of the mean, encompassing unobserved component gamma regression models for panel data. Models with serial correlation are briefly discussed.

1. Introduction

In the standard linear unobserved effects model, it is well known that consistent estimators are available under correct specification of the conditional mean and strict exogeneity of the explanatory variables, conditional on the latent individual effect. The usual fixed effects (within) estimator is consistent, as is the minimum chi-square estimator proposed by Chamberlain (1982).

To be more precise, let $\{(y_i, x_i, \phi_i) : i=1, 2, \dots\}$ be a sequence of independent, identically distributed random variables, where $y_i = (y_{i1}, \dots, y_{iT})'$ is $T \times 1$, $x_i = (x'_{i1}, \dots, x'_{iT})'$ is $T \times K$, and ϕ_i is the scalar unobserved effect. The linear unobserved effects model specifies that, for each $t=1, \dots, T$,

$$E(y_{it} | x_i, \phi_i) = E(y_{it} | x_{it}, \phi_i) = \phi_i + x_{it} \beta_o, \quad (1.1)$$

where β_o is a $K \times 1$ vector of unknown parameters. Equation (1.1) incorporates a linearity assumption and strict exogeneity of x_i conditional on the latent variable ϕ_i ; see Chamberlain (1984) for further discussion. Even though additional assumptions -- in particular, $V(y_i | x_i, \phi_i) = \sigma_o^2 I_T$ -- are typically imposed in carrying out inference after fixed effects estimation, assumption (1.1) and standard regularity conditions are sufficient for the fixed effects estimator to be consistent and asymptotically normally distributed. Thus, the fixed effects estimator is robust to conditional heteroskedasticity across individuals, as well as to serial correlation across time for a particular individual.

As noted by Chamberlain (1980), the fixed effects estimator is also the conditional maximum likelihood estimator (CMLE)¹ under the additional

assumption

$$y_i | x_i, \phi_i \sim N(\phi_i j_T + x_i \beta_0, \sigma_0^2 I_T), \quad (1.3)$$

where $j_T = (1, 1, \dots, 1)'$ is $T \times 1$. The fixed effects estimator of β_0 can be shown to maximize the log-likelihood based on the density of y_i conditional on $\sum_{t=1}^T y_{it}$, x_i , and ϕ_i , which turns out to be independent of ϕ_i . The consistency of the fixed effects estimator under (1.1) can therefore be interpreted as an important robustness property of the quasi-conditional maximum likelihood estimator (QCMLE) of β_0 .

Unlike linear models, little has been written on distribution-free estimation of nonlinear unobserved effects models. Chamberlain (1980, 1984) analyzes a fixed effects logit model, Hausman, Hall, and Griliches (1984) (hereafter, HHG) consider a variety of unobserved effects models for count panel data, and Papke (1989) estimates count models for firm births for a panel of states. All of these applications rely on the method of conditional maximum likelihood, where a sufficient statistic (the sum of the binary or count variable across time) is conditioned on to remove the unobserved effect. As far as I know, the robustness properties of these CMLEs to misspecification of the initially specified joint distribution have not been investigated. It is important to see that, even though the resulting conditional density (e.g. the multinomial) is typically in the linear exponential family (LEF), the robustness of the QCMLE in the LEF, e.g. Gouriéroux, Monfort, and Trognon (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 -- in particular models for count

data -- it would be useful to have a class of estimators that requires minimal distributional assumptions, while further relaxing the first two moment assumptions appearing in the literature. The conditional MLE approach is inherently limited by its reliance on a completely specified joint distribution. A new, distribution-free approach that nevertheless eliminates the unobserved effects for a broad class of models is needed. This paper develops the notion of a conditional linear predictor (CLP), and shows how CLPs can be used to eliminate individual effects in certain nonlinear unobserved effects models.

Section 2 introduces the model that motivated this research, and shows how the unobserved effects can be removed by computing an appropriate CLP. The conditional mean and variance assumptions are substantially more general than those implied by the most flexible negative binomial specification of HHG. In particular, the model allows not only for an individual effect in the mean, but it also allows for individual under- or overdispersion that can be unrelated to the mean effect. Independence across time is not assumed, and moments higher than the second are unrestricted.

Estimation of conditional linear predictors, which is a straightforward application of generalized method of moments (Hansen (1982)), is covered in section 3. Section 4 discusses specification testing in the context of CLPs. Section 5 analyzes the model of section 2 in detail, and suggests several consistent and asymptotically normal estimators. In particular, the multinomial QCMLE used by HHG for the fixed effects Poisson model is shown to be consistent and asymptotically normal much more generally.

Section 6 briefly covers a multiplicative unobserved effects model where the conditional variance is proportional to the square of the conditional mean,

as occurs in gamma and lognormal regression models. Section 7 outlines how serial correlation, conditional on the unobserved effects, can be accomodated.

2. Motivation: An Unobserved Effects Model for Count Panel Data

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

$$y_{it}|x_i, \phi_i \sim \text{Poisson}(\phi_i \mu(x_{it}, \beta_o)) \quad (2.1)$$

and

$$y_{it}, y_{ir} \text{ are independent conditional on } x_i, \phi_i, \quad t \neq r, \quad (2.2)$$

where

$$E(y_{it}|x_i, \phi_i) = E(y_{it}|x_{it}, \phi_i) = \phi_i \mu(x_{it}, \beta_o), \quad (2.3)$$

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

If a particular functional form for $E(\phi_i|x_i)$ is specified, then estimation of β_o can proceed under (2.3) only; further assumptions on $D(y_i|x_i, \phi_i)$ are not required. Equation (2.3), a model for $E(\phi_i|x_i)$, and the law of iterated expectations can be used to obtain $E(y_i|x_i)$ as a function of β_o and other parameters.² For example, if $\mu(x_{it}, \beta)$ is specified to be

$\exp(\mathbf{x}_{it}\beta)$, one might also assume that

$$E(\phi_i | \mathbf{x}_i) = \exp\left[\alpha_0 + \sum_{t=1}^T \mathbf{x}_{it} \lambda_{ot}\right]$$

for $K \times 1$ vectors λ_{ot} , $t=1, \dots, T$. However, in many cases one does not wish to be so precise about how ϕ_i and \mathbf{x}_i are related.

As an alternative to specifying $E(\phi_i | \mathbf{x}_i)$ (or $D(\phi_i | \mathbf{x}_i)$), HHG show how (2.1) and (2.2) can be used in Andersen's (1970) conditional ML methodology (see also Palmgren (1981) for the following derivation). Let $n_i = \sum_{t=1}^T y_{it}$ denote the sum across time of the explained variable. Then HHG show that

$$\mathbf{y}_i | n_i, \mathbf{x}_i, \phi_i \sim \text{Multinomial}(n_i, p_1(\mathbf{x}_i, \beta_0), \dots, p_T(\mathbf{x}_i, \beta_0)), \quad (2.4)$$

where

$$p_t(\mathbf{x}_i, \beta_0) = \mu(\mathbf{x}_{it}, \beta_0) / \left(\sum_{r=1}^T \mu(\mathbf{x}_{ir}, \beta_0) \right). \quad (2.5)$$

Because this distribution does not depend on ϕ_i , (2.4) is also the distribution of \mathbf{y}_i conditional on n_i and \mathbf{x}_i . Therefore, β_0 can be estimated by standard conditional MLE techniques. For later use, the conditional log-likelihood for observation i , apart from terms not depending on β , is

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

Because the multinomial distribution is in the LEF, the results of GMT imply a certain amount of robustness of the QCMLE. If (2.4) holds then

$$E(y_{it} | n_i, \mathbf{x}_i) = p_t(\mathbf{x}_i, \beta_0) n_i. \quad (2.7)$$

Conversely, it follows by GMT that, if (2.7) holds, then the 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 at least one other interesting case where (2.7) holds. Let α_i and γ_i be unobserved

individual effects. If

$$y_{it}|x_i, \alpha_i, \gamma_i \sim \text{Negative Binomial}(\alpha_i \mu(x_{it}, \beta_0), \gamma_i) \quad (2.8)$$

and

$$y_{it}, y_{ir} \text{ are independent conditional on } x_i, \alpha_i, \gamma_i, \quad t \neq r, \quad (2.9)$$

then (2.7) can be shown to hold. By GMT, the QCMLE based on the multinomial distribution provides consistent estimates of β_0 under (2.8) and (2.9). This is useful but still somewhat restrictive.

A robust approach consists of specifying at most a couple of low order conditional moments. Let ϕ_i and φ_i be scalar unobserved effects. A strictly weaker set of assumptions than (2.1)-(2.2) and (2.8)-(2.9) is

$$E(y_{it}|x_i, \phi_i, \varphi_i) = \phi_i \mu(x_{it}, \beta_0) \quad (2.10)$$

$$V(y_{it}|x_i, \phi_i, \varphi_i) = \varphi_i E(y_{it}|x_i, \phi_i, \varphi_i) = \varphi_i \phi_i \mu(x_{it}, \beta_0) \quad (2.11)$$

$$CV(y_{it}, y_{ir}|x_i, \phi_i, \varphi_i) = 0, \quad t \neq r. \quad (2.12)$$

Equations (2.10)-(2.12) specify the first two moments of y_i conditional on x_i and ϕ_i , and these are more general than the first two moments implied by (2.1)-(2.2) ($\varphi_i \equiv 1$) and as general as the first two moments implied by (2.8)-(2.9) ($\phi_i \equiv \alpha_i/\gamma_i$, $\varphi_i \equiv 1 + 1/\gamma_i$). Although (2.12) assumes zero conditional covariance, independence of the components of y_i conditional on x_i , ϕ_i , and φ_i is not assumed, nor is the distribution assumed to be Poisson, Negative Binomial, or anything in particular.

The primary question addressed in this paper is: In models such as (2.10)-(2.12), how can ϕ_i and φ_i be eliminated, so that β_0 can be estimated? One answer is really very simple. Define the sum of counts, n_i , as above. Then, as defined in section 3, the linear predictor of y_{it} on $(1, n_i)'$, conditional on (x_i, ϕ_i, φ_i) , is given by

$$\begin{aligned}
L(y_{it}|1, n_i; x_i, \phi_i, \varphi_i) &= E(y_{it}|x_i, \phi_i, \varphi_i) \\
&+ \frac{CV(y_{it}, n_i|x_i, \phi_i, \varphi_i)}{V(n_i|x_i, \phi_i, \varphi_i)} (n_i - E(n_i|x_i, \phi_i, \varphi_i)) \\
&- E(y_{it}|x_i, \phi_i, \varphi_i) \\
&+ \frac{V(y_{it}|x_i, \phi_i, \varphi_i)}{V(n_i|x_i, \phi_i, \varphi_i)} (n_i - E(n_i|x_i, \phi_i, \varphi_i)) \\
&= \phi_i \mu(x_{it}, \beta_o) + \frac{\phi_i \phi_i \mu(x_{it}, \beta_o)}{\sum_{r=1}^T \phi_i \phi_i \mu(x_{ir}, \beta_o)} (n_i - \sum_{r=1}^T \phi_i \mu(x_{ir}, \beta_o)) \\
&= \frac{\mu(x_{it}, \beta_o)}{\sum_{r=1}^T \mu(x_{ir}, \beta_o)} n_i. \tag{2.13}
\end{aligned}$$

There are a few points worth noting about this derivation. First, (2.13) is generally not the conditional expectation $E(y_{it}|n_i, x_i, \phi_i, \varphi_i)$, as was derived under (2.1) and (2.2) or (2.8) and (2.9). Thus, a class of estimators must be constructed to account for the fact that (2.13) represents $L(y_{it}|1, n_i; x_i, \phi_i, \varphi_i)$, but not necessarily $E(y_{it}|n_i, x_i, \phi_i, \varphi_i)$. Second, as is desired, this conditional linear predictor does not depend on ϕ_i or φ_i . Third, in this example, $L(y_{it}|1, n_i; x_i, \phi_i, \varphi_i) = L(y_{it}|n_i; x_i, \phi_i, \varphi_i)$, so that unity could have been excluded from the projection set. However, knowledge of this equality expands the type of orthogonality conditions that can be used in estimating β_o , and leads directly to the robustness result for the multinomial QCMLE.

I return to this example in section 5. The next two sections cover estimating and specification testing in the context of conditional linear

predictors.

3. Estimating Conditional Linear Predictors

This section defines and discusses estimation of conditional linear predictors. Unsurprisingly, intuition about linear predictors in an unconditional setting generally carries over to the conditional case. Let y be $J \times 1$, z be $K \times 1$, and w be $I \times 1$. In what follows, 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, unity can and should be included in z ; in the section 6 example, unity must be excluded from z .

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

$$\Sigma_{yz}(w) = E(yz' | w), \quad \Sigma_{zz}(w) = E(zz' | w). \quad (3.1)$$

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

DEFINITION 3.1: Let y , z , and w be defined as as above. The *linear predictor of y on z , conditional on w* , is defined to be

$$\begin{aligned} L(y|z;w) &= \Sigma_{yz}(w)\Sigma_{zz}^{-1}(w)z \\ &= C_o(w)z, \end{aligned} \quad (3.2)$$

where $C_o(w)$ is the $J \times K$ matrix

$$C_o(w) = \Sigma_{yz}(w)\Sigma_{zz}^{-1}(w). \quad \blacksquare$$

Note that $L(y|z;w)$ is always linear in z , but is generally a nonlinear function of w . When the context is clear, $L(y|z;w)$ is simply called a

conditional linear predictor (CLP). The difference between y and its CLP has zero orthogonality properties that are immediate extensions from unconditional linear predictor theory.

LEMMA 3.1: Let y , z , and w be as in Definition 3.1. Define

$$u \equiv y - L(y|z;w) = y - C_0(w)z. \quad (3.3)$$

Then

$$E(uz' | w) = 0. \quad (3.4)$$

PROOF: $uz' = [y - C_0(w)z]z' = yz' - C_0(w)zz'$, so that

$$\begin{aligned} E(uz' | w) &= E(yz' | w) - C_0(w)E(zz' | w) \\ &= \Sigma_{yz}(w) - \Sigma_{yz}(w)\Sigma_{zz}^{-1}(w)\Sigma_{zz}(w) \\ &= 0. \quad \blacksquare \end{aligned}$$

The next corollary, which motivates the class of estimators considered, follows immediately by the law of iterated expectations.

COROLLARY 3.1: Let y , z , w , and u be as in Lemma 3.1, and let $D(w)$ be a $JK \times L$ random matrix. Then

$$E[D(w)'(z \otimes I_J)u] = E[D(w)'\text{vec}\{uz'\}] = 0. \quad \blacksquare \quad (3.5)$$

Suppose now that $C_0(w) \equiv C(w, \theta_0)$, where $C(w, \theta)$ is a known function of w and the $P \times 1$ parameter vector $\theta \in \Theta$. Then, for a matrix function $D(w)$ as defined in Corollary 3.1, θ_0 solves (perhaps not uniquely) the system of equations

$$E[D(w)'(z \otimes I_J)\{y - C(w, \theta)\}] = 0. \quad (3.6)$$

Equation (3.6) can be exploited to obtain a variety of consistent estimators of θ_0 .

For the remainder of this section, let $\{(y_i, z_i, w_i) : i=1, 2, \dots\}$ be an

i.i.d sequence, where y_i is $J \times 1$, z_i is $K \times 1$, and w_i is 1×1 . Extension of the subsequent results to heterogeneous and/or dependent situations is fairly straightforward but notationally cumbersome. The available sample size is denoted N .

Assume that for a known function $C(w_i, \theta)$,

$$L(y_i | z_i; w_i) = C(w_i, \theta_0) z_i. \quad (3.7)$$

In the applications, not all of the vector w_i is observed ($w_i = (x_i, \phi_i, \varphi_i)$ in (2.10)-(2.12)), and $C(w_i, \theta_0)$ does not depend on the unobserved elements. For notational simplicity, this section treats w_i as entirely observed. When w_i contains unobservables, the orthogonality conditions constructed below are necessarily restricted to functions of the observables.

The class of estimators is assumed to solve a first order condition asymptotically. To specify the estimating equations, let $D(w_i, \theta, \gamma)$ be a $JK \times P$ matrix depending on w_i , θ , and possibly a vector of nuisance parameters, $\gamma \in \Gamma$. Assume that an estimator $\hat{\gamma}_N$ is available such that

$$\sqrt{N}(\hat{\gamma}_N - \gamma^*) = o_p(1) \text{ for some } \gamma^* \in \Gamma. \quad (3.8)$$

Then, $\hat{\theta}_N$ is assumed to satisfy

$$N^{-1/2} \sum_{i=1}^N D(w_i, \hat{\theta}_N, \hat{\gamma}_N)' [z_i \otimes I_J] \{y_i - C(w_i, \hat{\theta}_N) z_i\} = o_p(1); \quad (3.9)$$

in shorthand,

$$N^{-1/2} \sum_{i=1}^N D_i(\hat{\theta}_N, \hat{\gamma}_N)' [z_i \otimes I_J] u_i(\hat{\theta}_N) = o_p(1),$$

where $u_i(\theta) = y_i - C(w_i, \theta) z_i$. As further shorthand, let $u_i = u_i(\theta_0)$, $\hat{D}_i = D_i(\hat{\theta}_N, \hat{\gamma}_N)$, and $\hat{u}_i = u_i(\hat{\theta}_N)$. In all of the examples in this paper, $\hat{\theta}_N$ is an exact solution to the P equations

$$\sum_{i=1}^N D_i(\theta, \hat{\gamma}_N)' [z_i \otimes I_J] u_i(\theta) = 0, \quad (3.10)$$

so that (3.9) is trivially satisfied.

The weak consistency of $\hat{\theta}_N$ for θ_0 hinges on a standard uniform weak law of large numbers and a suitable identification condition. Identification requires that θ_0 is the only element of

$$\mathcal{S} = \{\theta \in \Theta: E[D_i(\theta, \gamma^*)' (z_i \otimes I_J) u_i(\theta)] = 0\}. \quad (3.11)$$

By Corollary 3.1, θ_0 is an element of \mathcal{S} ; as usual, this must be strengthened to the assumption that θ_0 is the unique solution to the asymptotic orthogonality condition.

Establishing the asymptotic normality of $\sqrt{N}(\hat{\theta}_N - \theta_0)$ is also relatively straightforward, but a little algebraic care is required to show that the natural estimate of the asymptotic variance matrix of $\sqrt{N}(\hat{\theta}_N - \theta_0)$ is valid. The slight complication arises because $E(u_i | w_i, z_i) \neq 0$ necessarily; (3.7) guarantees only that $E(u_i z_i' | w_i) = 0$.

The first step in deriving the asymptotic distribution of $\sqrt{N}(\hat{\theta}_N - \theta_0)$ is standard; it amounts to showing that the asymptotic distribution of $\sqrt{N}(\hat{\theta}_N - \theta_0)$ does not depend on that of $\sqrt{N}(\hat{\gamma}_N - \gamma^*)$. This follows by a mean value expansion:

$$\begin{aligned} N^{-1/2} \sum_{i=1}^N D_i(\hat{\theta}_N, \hat{\gamma}_N)' [z_i \otimes I_J] u_i(\hat{\theta}_N) &= N^{-1/2} \sum_{i=1}^N D_i(\hat{\theta}_N, \gamma^*)' [z_i \otimes I_J] u_i(\hat{\theta}_N) \\ &+ E[(u_i' (z_i' \otimes I_J) \otimes I_p) \partial \text{vec}(D_i(\theta_0, \gamma^*)') / \partial \gamma] \sqrt{N}(\hat{\gamma}_N - \gamma^*) + o_p(1). \end{aligned}$$

But

$$E[(u_i' (z_i' \otimes I_J) \otimes I_p) | w_i] = 0,$$

so that

$$\begin{aligned}
& N^{-1/2} \sum_{i=1}^N D_i(\hat{\theta}_N, \hat{\gamma}_N)' [z_i \otimes I_J] u_i(\hat{\theta}_N) \\
& = N^{-1/2} \sum_{i=1}^N D_i(\hat{\theta}_N, \gamma^*)' [z_i \otimes I_J] u_i(\hat{\theta}_N) + o_p(1).
\end{aligned} \tag{3.12}$$

Therefore, define the $P \times 1$ "score" vector

$$s_i(\theta) = D_i(\theta, \gamma^*)' [z_i \otimes I_J] u_i(\theta), \tag{3.13}$$

so that $E[s_i(\theta_o) | w_i] = 0$ by Corollary 3.1. Another mean value expansion gives

$$o_p(1) = N^{-1/2} \sum_{i=1}^N s_i(\theta_o) + E[H_i(\theta_o)] \sqrt{N}(\hat{\theta}_N - \theta_o),$$

where $H_i(\theta) = \nabla_{\theta} s_i(\theta)$ is the $P \times P$ derivative of $s_i(\theta)$. Provided that

$$A_o = -E[H_i(\theta_o)] \text{ is nonsingular,} \tag{3.14}$$

$\sqrt{N}(\hat{\theta}_N - \theta_o)$ has a familiar asymptotic representation:

$$\sqrt{N}(\hat{\theta}_N - \theta_o) = A_o^{-1} N^{-1/2} \sum_{i=1}^N s_i(\theta_o) + o_p(1).$$

Letting

$$\begin{aligned}
B_o &= E[s_i(\theta_o) s_i(\theta_o)'] \\
&= E[D_i(\theta_o, \gamma^*)' [z_i \otimes I_J] u_i u_i' [z_i' \otimes I_J] D_i(\theta_o, \gamma^*)],
\end{aligned} \tag{3.15}$$

it follows that

$$\sqrt{N}(\hat{\theta}_N - \theta_o) \overset{d}{\rightarrow} N(0, A_o^{-1} B_o A_o^{-1}). \tag{3.16}$$

This discussion is summarized with an informal theorem.

THEOREM 3.1: Under (3.7), (3.8), (3.9), (3.11), (3.14), and standard regularity conditions, (3.16) holds. ■

The matrix B_o is easily estimated by a standard outer product of the "score":

$$\hat{B}_N = N^{-1} \sum_{i=1}^N s_i(\hat{\theta}_N, \hat{\gamma}_N) s_i(\hat{\theta}_N, \hat{\gamma}_N)' \quad (3.17)$$

$$\begin{aligned} &= N^{-1} \sum_{i=1}^N D_i(\hat{\theta}_N, \hat{\gamma}_N)' [z_i \otimes I_J] u_i(\hat{\theta}_N) u_i(\hat{\theta}_N)' [z_i' \otimes I_J] D_i(\hat{\theta}_N, \hat{\gamma}_N) \\ &= N^{-1} \sum_{i=1}^N \hat{D}_i' [z_i \otimes I_J] \hat{u}_i \hat{u}_i' [z_i' \otimes I_J] \hat{D}_i, \end{aligned} \quad (3.18)$$

which is at least positive semi-definite. The most convenient estimator of A_o excludes terms depending on the derivative of $D_i(\theta, \gamma^*)$ with respect to θ . But

$$\begin{aligned} H_i(\theta) &= -D_i(\theta, \gamma^*)' [z_i \otimes I_J] [z_i' \otimes I_J] \nabla_{\theta} C_i(\theta) \\ &\quad + \{u_i(\theta)' \otimes I_P\} \{ (z_i' \otimes I_J) \otimes I_P \} \partial \text{vec}[D_i(\theta, \gamma^*)'] / \partial \theta, \end{aligned}$$

where $\nabla_{\theta} C_i(\theta) \equiv \partial \text{vec}[C_i(\theta)] / \partial \theta$ is $JK \times P$ and $\partial \text{vec}[D_i(\theta, \gamma^*)'] / \partial \theta$ is $JKP \times P$.

Therefore,

$$\begin{aligned} -H_i(\theta_o) &= D_i(\theta_o, \gamma^*)' [z_i \otimes I_J] [z_i' \otimes I_J] \nabla_{\theta} C_i(\theta_o) \\ &\quad - \{u_i(\theta_o)' (z_i' \otimes I_J) \otimes I_P\} \partial \text{vec}[D_i(\theta_o, \gamma^*)'] / \partial \theta. \end{aligned} \quad (3.19)$$

Under (3.7),

$$E[u_i(\theta_o)' (z_i' \otimes I_J) | w_i] = 0$$

and, because $\partial \text{vec}[D_i(\theta_o, \gamma^*)'] / \partial \theta$ depends only on w_i ,

$$A_o = E[D_i(\theta_o, \gamma^*)' [z_i z_i' \otimes I_J] \nabla_{\theta} C_i(\theta_o)]. \quad (3.20)$$

It follows that a consistent estimator of A_o is simply

$$\hat{A}_N = N^{-1} \sum_{i=1}^N \hat{D}_i' [z_i z_i' \otimes I_J] \nabla_{\theta} \hat{C}_i. \quad (3.21)$$

Inference is carried out on θ_0 by treating $\hat{\theta}_N$ as normally distributed with mean θ_0 and variance

$$\hat{A}_N^{-1} \hat{B}_N \hat{A}_N^{-1} / N. \quad (3.22)$$

Several special cases can be cast in terms of (3.9). One useful class of estimators is multivariate weighted nonlinear least squares (MWNLS). Given a $J \times J$ symmetric, positive semidefinite matrix $G(w_i, \hat{\gamma}_N)$, choose $\hat{\theta}_N$ to solve

$$\min_{\theta \in \Theta} \sum_{i=1}^N [y_i - C(w_i, \theta) z_i]' G(w_i, \hat{\gamma}_N) [y_i - C(w_i, \theta) z_i]. \quad (3.23)$$

Note that the weighting matrix $G(w_i, \gamma)$ is allowed to depend on w_i but not on z_i . Here, $\hat{\gamma}_N$ is an initial estimator. The MWNLS estimator falls under Theorem 3.1 by choosing

$$D(w_i, \theta, \gamma)' = [\nabla_{\theta} C_1(w_i, \theta)' G(w_i, \gamma) \mid \cdots \mid \nabla_{\theta} C_K(w_i, \theta)' G(w_i, \gamma)], \quad (3.24)$$

where $\nabla_{\theta} C(w_i, \theta)$ has been partitioned as

$$\nabla_{\theta} C(w_i, \theta)' = [\nabla_{\theta} C_1(w_i, \theta)' \mid \cdots \mid \nabla_{\theta} C_K(w_i, \theta)'],$$

and $\nabla_{\theta} C_k(w_i, \theta)$ is $J \times P$, $k=1, \dots, K$. Then

$$\begin{aligned} D(w_i, \theta, \gamma)' [z_i \otimes I_J] u_i(\theta) &= \nabla_{\theta} C(w_i, \theta)' [z_i \otimes G(w_i, \gamma)] u_i(\theta) \\ &= \nabla_{\theta} C(w_i, \theta)' [z_i \otimes I_J] G(w_i, \gamma) u_i(\theta). \end{aligned}$$

In terms of more familiar notation, let

$$m_i(\theta) \equiv m(w_i, z_i, \theta) \equiv C(w_i, \theta) z_i \quad (3.25)$$

denote the "regression" function (but recall that $m(w_i, z_i, \theta_0) \neq E(y_i | w_i, z_i)$ necessarily). Then

$$M_i(\theta) = \nabla_{\theta} m_i(\theta) - [z_i' \otimes I_J] \nabla_{\theta} G_i(\theta), \quad (3.26)$$

so that

$$s_i(\theta_0) = M_i(\theta_0)' G_i(\gamma^*) u_i \quad (3.27)$$

and

$$A_0 = E[M_i(\theta_0)' G_i(\gamma^*) M_i(\theta_0)]. \quad (3.28)$$

The consistent estimators of A_0 and B_0 are simply

$$\hat{A}_N = N^{-1} \sum_{i=1}^N \nabla_{\theta} \hat{G}_i' [z_i \otimes I_J] \hat{G}_i [z_i' \otimes I_J] \nabla_{\theta} \hat{G}_i = N^{-1} \sum_{i=1}^N \hat{M}_i' \hat{G}_i \hat{M}_i \quad (3.29)$$

and

$$\hat{B}_N = N^{-1} \sum_{i=1}^N \hat{M}_i' \hat{G}_i \hat{u}_i \hat{u}_i' \hat{G}_i \hat{M}_i = N^{-1} \sum_{i=1}^N \hat{s}_i \hat{s}_i', \quad (3.30)$$

which are the familiar robust formulas from MWNLS theory when estimating conditional expectations. Thus, even when estimating conditional linear predictors, simple positive definite estimates of A_0 and B_0 are available.

Only rarely does it happen that

$$E(u_i u_i' | w_i, z_i) = [G(w_i, \gamma^*)]^{-1}, \quad (3.31)$$

in which case $A_0 = B_0$ and either \hat{A}_N^{-1}/N or \hat{B}_N^{-1}/N can be used to estimate the asymptotic variance of $\hat{\theta}_N$.

The generalized method of moments estimators studied by Hansen (1982) are also covered by Theorem 3.1. Let $L(w_i, \phi)$ be a $JK \times M$ matrix depending on w_i and a nuisance parameter ϕ . Assume that $\sqrt{N}(\hat{\phi}_N - \phi^*) = O_p(1)$. Let $\hat{\Xi}_N$ denote a symmetric, positive semi-definite matrix estimator such that $\sqrt{N}(\hat{\Xi}_N - \Xi^*) = O_p(1)$, where Ξ^* is a symmetric, positive definite matrix. In the current context, a GMM estimator $\hat{\theta}_N$ solves

$$\min_{\theta \in \Theta} \tau_N(\theta)$$

where

$$r_N(\theta) = \left[N^{-1} \sum_{i=1}^N L(w_i, \hat{\phi}_N)' [z_i \otimes I_J] u_i(\theta) \right]' \hat{\Xi}_N \left[N^{-1} \sum_{i=1}^N L(w_i, \hat{\phi}_N)' [z_i \otimes I_J] u_i(\theta) \right]. \quad (3.32)$$

Under differentiability assumptions, this estimator can be shown to be a special case of Theorem 3.1. The first order condition solved by $\hat{\theta}_N$ is

$$\left[\sum_{i=1}^N \hat{L}'_i [z_i z_i' \otimes I_J] \nabla_{\theta} C_i(\theta) \right]' \hat{\Xi}_N \left[\sum_{i=1}^N \hat{L}'_i [z_i \otimes I_J] u_i(\theta) \right] = 0.$$

Letting

$$\hat{R}_N = \left[N^{-1} \sum_{i=1}^N L_i(\hat{\phi}_N)' [z_i z_i' \otimes I_J] \nabla_{\theta} C_i(\hat{\theta}_N) \right], \quad (3.33)$$

which is an $M \times P$ matrix, $\hat{\theta}_N$ equivalently solves

$$N^{-1/2} \sum_{i=1}^N \hat{R}'_N \hat{\Xi}_N L(w_i, \hat{\phi}_N)' [z_i \otimes I_J] u_i(\theta) = 0. \quad (3.34)$$

Therefore, in the notation of Theorem 3.1, let

$$D(w_i, \hat{\gamma}_N) = L(w_i, \hat{\phi}_N) \hat{\Xi}_N \hat{R}_N', \quad (3.35)$$

where $\hat{\gamma}_N = (\hat{\phi}'_N, \{\text{vech}(\hat{\Xi}_N)\}', \{\text{vec}(\hat{R}_N)\}')'$; this choice of D does not depend on θ .

Although the GMM estimator is consistent and asymptotically normal for a variety of weighting matrices $\hat{\Xi}_N$, the efficient estimator -- given the choice of $L(w_i, \phi^*)$ -- is always available. This is the minimum chi-square estimator, obtained by choosing $\hat{\Xi}_N$ to be a consistent estimator of

$$[E[L(w_i, \phi^*)' [z_i \otimes I_J] u_i u_i' [z_i' \otimes I_J] L(w_i, \phi^*)]]^{-1}. \quad (3.36)$$

Subsequently, $\hat{\Xi}_N$ is assumed to be chosen in this way. This requires an initial, consistent estimator of θ_0 , such as a MWNLS estimator.

The asymptotic variance of $\sqrt{N}(\hat{\theta}_N - \theta_0)$, where $\hat{\theta}_N$ now denotes the minimum chi-square estimator, is $A_0^{-1} - B_0^{-1}$, because A_0 and B_0 are both equal to $R^* \Xi^* R^*$, where Ξ^* is given by (3.36) and

$$R^* = E[L(w_i, \phi^*)' [z_i z_i' \otimes I_J] \nabla_{\theta} C(w_i, \theta_0)].$$

The asymptotic variance of $\hat{\theta}_N$ is estimated by $(\hat{R}_N' \hat{\Xi}_N \hat{R}_N)^{-1}/N$.

The problem of estimating conditional linear predictors also fits into the framework of Chamberlain (1987), who derives the efficiency bound for estimators derived from conditional moment restrictions. For CLPs, the conditional moment restrictions available for estimating θ_0 are given by

$$E[(y_i - C(w_i, \theta_0)z_i)z_i' | w_i] = 0$$

or, in vector notation,

$$E[(z_i \otimes I_J)(y_i - C(w_i, \theta_0)z_i) | w_i] = 0$$

(a total of JK conditional moment restrictions). Letting

$$\Sigma_0(w_i) = E[(z_i \otimes I_J)u_i u_i' (z_i' \otimes I_J') | w_i]$$

and

$$\Psi_0(w_i) = E[(z_i z_i' \otimes I_J) | w_i] \nabla_{\theta} C(w_i, \theta_0),$$

the lower bound is obtained from Chamberlain (1987, equation (1.11)):

$$(E[\Psi_0(w_i)' \Sigma_0^{-1}(w_i) \Psi_0(w_i)])^{-1}.$$

To achieve this bound, one can proceed as in Newey (1987) and nonparametrically estimate $\Sigma_0(w_i)$ and $E[(z_i z_i' \otimes I_J) | w_i]$. Although studying this kind of procedure is beyond the scope of this paper, the lower bound calculation at least isolates which terms need to be nonparametrically estimated.

4. Specification Testing

Specification tests can be derived using the approach of Newey (1985).

If $\Lambda(w_i, \theta, \gamma)$ is a $JK \times Q$ matrix depending on w_i , θ , and nuisance parameters γ , a general class of tests is based on the sample covariance

$$N^{-1} \sum_{i=1}^N \Lambda(w_i, \hat{\theta}_N, \hat{\gamma}_N)' [z_i \otimes I_J] u_i(\hat{\theta}_N). \quad (4.1)$$

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

$$\begin{aligned} & N^{-1/2} \sum_{i=1}^N \Lambda(w_i, \hat{\theta}_N, \hat{\gamma}_N)' [z_i \otimes I_J] u_i(\hat{\theta}_N) \\ &= N^{-1/2} \sum_{i=1}^N \Lambda(w_i, \hat{\theta}_N, \gamma^*)' [z_i \otimes I_J] u_i(\hat{\theta}_N) + o_p(1) \\ &= N^{-1/2} \sum_{i=1}^N \Lambda(w_i, \theta_o, \gamma^*)' [z_i \otimes I_J] u_i \\ &\quad - E[\Lambda_i(\theta_o, \gamma^*)' [z_i z_i' \otimes I_J] \nabla_{\theta} C_i(\theta_o)] \sqrt{N}(\hat{\theta}_N - \theta_o) + o_p(1) \\ &\equiv N^{-1/2} \sum_{i=1}^N \varphi_i(\theta_o, \gamma^*) - K_o' A_o^{-1} N^{-1/2} \sum_{i=1}^N s_i(\theta_o, \gamma^*) + o_p(1) \\ &= N^{-1/2} \sum_{i=1}^N \{\Lambda_i(\theta_o, \gamma^*) - D_i(\theta_o, \gamma^*) A_o^{-1} K_o\}' [z_i \otimes I_J] u_i + o_p(1) \end{aligned}$$

where A_o is given by (3.20),

$$\varphi_i(\theta_o, \gamma^*) \equiv \Lambda(w_i, \theta_o, \gamma^*)' [z_i \otimes I_J] u_i,$$

and

$$K_o \equiv E[\nabla_{\theta} C_i(\theta_o)' [z_i z_i' \otimes I_J] \Lambda_i(\theta_o, \gamma^*)].$$

Thus, let

$$\hat{K}_N \equiv N^{-1} \sum_{i=1}^N \nabla_{\theta} C_i(\hat{\theta}_N)' [z_i z_i' \otimes I_J] \Lambda_i(\hat{\theta}_N, \hat{\gamma}_N) \quad (4.2)$$

and

$$\hat{\xi}'_i = (\hat{\Lambda}_i - \hat{D}_i \hat{A}_N^{-1} \hat{K}_N)' [z_i \otimes I_j] \hat{u}_i, \quad i=1, \dots, N \quad (4.3)$$

(note that $\hat{\xi}_i$ is a $1 \times Q$ vector). A valid test statistic is obtained as $NR_u^2 = N - \text{SSR}$ from the regression

$$1 \text{ on } \hat{\xi}_i, \quad i=1, \dots, N. \quad (4.4)$$

Under the null hypothesis

$$H_0: E[u_i(\theta_0) z_i' | w_i] = 0, \quad (4.5)$$

$N - \text{SSR} \stackrel{a}{\sim} \chi_Q^2$, provided there are no redundant columns in $\Lambda_i(\theta_0, \gamma^*)$. As a special case, this procedure covers a robust, regression-based Hausman test for comparing the multinomial QCMLE and MNLS in the nonlinear unobserved effects model (2.10)-(2.12). This test is discussed in detail in section 5.

If the minimum chi-square estimator is used, where the number of orthogonality conditions M is greater than the number of parameters P , then the GMM overidentification test is available from Hansen (1982). The test statistic is simply N times the value of the minimum chi-square objective function. Under (4.5),

$$N\tau_N(\hat{\theta}_N) \stackrel{a}{\sim} \chi_{M-P}^2, \quad (4.6)$$

where $\tau_N(\theta)$ is defined in (3.32) with appropriate choice of $\hat{\Xi}_N$.

5. Application to Count Models with Individual-Specific Dispersion

This section applies the theory of sections 3 and 4 to the model introduced in section 2. Nothing of what follows relies on y_i being a vector of counts, but the example is motivated by the count models of HHG. Let $\{(y_i, x_i, \phi_i, \varphi_i): i=1, 2, \dots\}$ be a sequence of i.i.d. random variables. As in

section 2, y_i is $T \times 1$, x_i is $T \times K$, and these are observed. ϕ_i and φ_i are unobserved, scalar random variables representing the individual effects. For clarity, the model introduced in section 2 is reproduced here. For $t=1, \dots, T$,

$$E(y_{it}|x_i, \phi_i, \varphi_i) = \phi_i \mu(x_{it}, \beta_0) \quad (5.1)$$

$$V(y_{it}|x_i, \phi_i, \varphi_i) = \varphi_i E(y_{it}|x_i, \phi_i, \varphi_i) = \varphi_i \phi_i \mu(x_{it}, \beta_0) \quad (5.2)$$

$$CV(y_{it}, y_{ir}|x_i, \phi_i, \varphi_i) = 0, \quad t \neq r. \quad (5.3)$$

This model allows for individual mean effects as well as a separate, individual dispersion, with variance to mean ratio

$$V(y_{it}|x_i, \phi_i, \varphi_i)/E(y_{it}|x_i, \phi_i, \varphi_i) = \varphi_i. \quad (5.4)$$

The addition of φ_i allows for under- or overdispersion, depending on the individual. Assumptions (5.1), (5.2), and (5.3) are more flexible than the first two moments of all of the fixed effects models used by HHG. The FEP model imposes $\varphi_i = 1$. The fixed effects negative binomial (FENB) model of HHG imposes $\varphi_i = 1 + \phi_i$. Not only is underdispersion ruled out for all individuals, but the amount of overdispersion is tied directly to the mean effect.⁴ In addition, (5.3) is weaker than independence, and no distributional assumption is made.

Section 2 showed that the linear predictor of y_{it} on $(1, n_i)'$, conditional on (x_i, ϕ_i, φ_i) , is free of ϕ_i and φ_i . In vector notation,

$$L(y_i|1, n_i; x_i, \phi_i, \varphi_i) = p(x_i, \beta_0) n_i, \quad (5.5)$$

where $p(x_i, \beta)$ denotes the $T \times 1$ vector with t^{th} element

$$p_t(x_i, \beta) = \frac{\mu(x_{it}, \beta)}{\sum_{r=1}^T \mu(x_{ir}, \beta)}. \quad (5.6)$$

Equation (5.5) implies orthogonality conditions of the form

$$E[D(x_i)' \{(1, n_i)' \otimes I_T\} u_i] = 0, \quad (5.7)$$

where $D(x_i)$ is any $2T \times L$ matrix function of x_i , and $u_i = y_i - p(x_i, \beta_0)n_i$.

This allows for a variety of method of moments procedures, as well as some simple, well known estimators. For example, Theorem 3.1 implies that the

MNLS estimator $\hat{\beta}_N$, which solves

$$\min_{\beta \in B} \sum_{i=1}^N (y_i - p(x_i, \beta)n_i)' (y_i - p(x_i, \beta)n_i), \quad (5.8)$$

is consistent for β_0 and asymptotically normally distributed. The asymptotic variance of $\sqrt{N}(\hat{\beta}_N - \beta_0)$ is $A_0^{-1} B_0 A_0^{-1}$, and consistent estimators of A_0 and B_0 are given by

$$\hat{A}_N = N^{-1} \sum_{i=1}^N n_i^2 \nabla_{\beta} p_i' \nabla_{\beta} p_i \quad (5.9)$$

and

$$\hat{B}_N = N^{-1} \sum_{i=1}^N n_i^2 \nabla_{\beta} p_i' \hat{u}_i \hat{u}_i' \nabla_{\beta} p_i, \quad (5.10)$$

where $\nabla_{\beta} p_i = \nabla_{\beta} p(x_i, \hat{\beta}_N)$ is the TxP gradient of $p(x_i, \beta)$ evaluated at $\hat{\beta}_N$. This is easily extended to MWNLS with TxT weighting matrix $\hat{G}_i = G(x_i, \hat{\gamma}_N)$. The MWNLS estimator solves

$$\min_{\beta \in B} \sum_{i=1}^N (y_i - p(x_i, \beta)n_i)' \hat{G}_i (y_i - p(x_i, \beta)n_i), \quad (5.11)$$

and A_0 and B_0 are easily estimated by

$$\hat{A}_N = N^{-1} \sum_{i=1}^N n_i^2 \nabla_{\beta} p_i' \hat{G}_i \nabla_{\beta} p_i \quad (5.12)$$

and

$$\hat{B}_N = N^{-1} \sum_{i=1}^N n_i^2 \nabla_{\beta} \hat{p}_i' \hat{G}_i \hat{u}_i \hat{u}_i' \hat{G}_i \nabla_{\beta} \hat{p}_i. \quad (5.13)$$

Even more interesting is that the multinomial QCMLE (or FEP QCMLE) is consistent under only assumptions (5.1)-(5.3). This is remarkable given that the conditional multinomial distribution was derived from a Poisson distribution with only one unobserved effect, i.e. $\varphi_i = 1$ (see section 2). Moreover, the distribution can be very different from the Poisson, and independence of the elements of y_i conditional on (x_i, ϕ_i, φ_i) is not assumed.

To see that the multinomial QCMLE is covered by Theorem 3.1, note that the gradient of the quasi-log likelihood (see (2.6)) is

$$\begin{aligned} s_i(\beta) &= \nabla_{\beta} \ell_i(\beta) = \sum_{t=1}^T y_{it} [\nabla_{\beta} p_t(x_i, \beta)' / p_t(x_i, \beta)] \\ &= \sum_{t=1}^T [\nabla_{\beta} p_t(x_i, \beta)' / p_t(x_i, \beta)] (y_{it} - p_t(x_i, \beta) n_i) \end{aligned} \quad (5.14)$$

$$\begin{aligned} &= \nabla_{\beta} p(x_i, \beta)' W(x_i, \beta) (y_i - p(x_i, \beta) n_i) \\ &= \nabla_{\beta} p(x_i, \beta)' W(x_i, \beta) u_i(\beta), \end{aligned} \quad (5.15)$$

where $W(x_i, \beta) = [\text{diag}(p_1(x_i, \beta), \dots, p_T(x_i, \beta))]^{-1}$. Equation (5.14) follows

from the fact that $\sum_{t=1}^T p_t(x_i, \beta) = 1$ for all β . Because $z_i' = (1, n_i)$, (5.15)

is seen to be of the form (3.10) with $D_i(\beta)' = [\nabla_{\beta} p_i(\beta)' W_i(\beta) \mid 0]$. If the FEP model is maintained, the estimate of the asymptotic variance of $\hat{\beta}_N$, obtained from the estimated information matrix, is

$$\begin{aligned}\hat{A}_N^{-1}/N &= \left[\sum_{i=1}^N n_i \nabla_{\beta} p(x_i, \hat{\beta}_N)' W(x_i, \hat{\beta}_N) \nabla_{\beta} p(x_i, \hat{\beta}_N) \right]^{-1} \\ &= \left[\sum_{i=1}^N \nabla_{\beta} m_i' \hat{V}_i^{-1} \nabla_{\beta} m_i \right]^{-1},\end{aligned}\quad (5.16)$$

where $m_i(\beta) = p_i(\beta)n_i$ would normally be the conditional mean function associated with the multinomial distribution, $\nabla_{\beta} m_i = \nabla_{\beta} m_i(\hat{\beta}_N) = \nabla_{\beta} p_i(\hat{\beta}_N)n_i$, and $\hat{V}_i = V(n_i, x_i, \hat{\beta}_N) = \text{diag}(p_{i1}(\hat{\beta}_N)n_i, \dots, p_{iT}(\hat{\beta}_N)n_i)$. Expression (5.16) is familiar from standard likelihood theory involving the multinomial distribution.

Unless the original Poisson model holds, \hat{A}_N^{-1}/N produces inappropriate standard errors. The robust form is $\hat{A}_N^{-1} \hat{B}_N \hat{A}_N^{-1}/N$, where

$$\begin{aligned}\hat{B}_N &= N^{-1} \sum_{i=1}^N \nabla_{\beta} p(x_i, \hat{\beta}_N)' W(x_i, \hat{\beta}_N) \hat{u}_i \hat{u}_i' W(x_i, \hat{\beta}_N) \nabla_{\beta} p(x_i, \hat{\beta}_N) \\ &= N^{-1} \sum_{i=1}^N \nabla_{\beta} m_i' \hat{V}_i^{-1} \hat{u}_i \hat{u}_i' \hat{V}_i^{-1} \nabla_{\beta} m_i.\end{aligned}\quad (5.17)$$

The estimator $\hat{A}_N^{-1} \hat{B}_N \hat{A}_N^{-1}/N$ is robust to arbitrary serial correlation in $\{u_{it} : t=1, 2, \dots, T\}$. Note that, by definition, $\sum_{t=1}^T u_{it} = 0$, so that the u_{it} might generally be expected to exhibit negative serial correlation. This is the case under (2.1) and (2.2). From McCullagh and Nelder (1989, p.165) the correlation between u_{it} and u_{ir} , conditional on (n_i, x_i) , is

$$-p_{it}(\beta_o)p_{ir}(\beta_o)/[p_{it}(\beta_o)(1-p_{it}(\beta_o))p_{ir}(\beta_o)(1-p_{ir}(\beta_o))]^{1/2}. \quad (5.18)$$

This particular negative correlation, which is used implicitly in the estimator \hat{A}_N^{-1}/N of the asymptotic variance of $\hat{\beta}_N$, need no longer hold under (5.1)-(5.3). In fact, it is no longer possible to compute the correlation

between u_{it} and u_{ir} , conditional on (n_i, x_i) , under assumptions (5.1)-(5.3). Thus, the robust covariance matrix estimator should always be computed; this can produce standard errors smaller or larger than those obtained from (5.16).

The robustness of the QCMLE to distributional misspecification suggests a research methodology different from that used by HHG. They compute a specification test for the FEP model that checks whether the data support the serial correlation pattern (5.18). HHG properly view a violation of (5.18) as a rejection of the original Poisson specification. However, this section has shown that the estimates of β_0 in model (5.1)-(5.3) are still consistent and asymptotically normal, whether or not the correlation structure (5.18) holds. Rather than testing whether the multinomial QCMLE estimates are consistent for β_0 , the HHG test looks for departures from the multinomial distributional assumption. A test of model (5.1)-(5.3) should be based on the testable implication that the linear predictor of y_i on $(1, n_i)'$, conditional on (x_i, ϕ_i, φ_i) , is of the form (5.5). Because QCMLE and MNLS are both consistent for β_0 under (5.5), a robust form of Hausman's (1978) test for comparing the two estimators is natural. Here I focus on a regression form of the test that requires computation of the QCMLE only, and results in a particularly simple research methodology.

The regression-based Hausman test is a special case of the tests discussed in section 4, but it is more directly obtained from Wooldridge (1991). Because the Poisson model is the nominal distribution for count data, it makes sense to construct the robust test to be optimal if the Poisson model is true, and the tests in Wooldridge (1991) are constructed in this manner. Let \hat{u}_i , $\nabla_{\beta} \hat{m}_i$, and \hat{V}_i be defined as above, evaluated at the QCMLE $\hat{\beta}_N$.

Define the weighted quantities $\tilde{u}_i = \hat{V}_i^{-1/2} \hat{u}_i$, $\nabla_{\beta} \tilde{m}_i = \hat{V}_i^{-1/2} \nabla_{\beta} \hat{m}_i$, and $\tilde{\lambda}_i = \hat{V}_i^{1/2} \nabla_{\beta} \hat{m}_i$ (this lambda is related, but not equal, to that appearing in section 4). The robust Hausman test is easily computed by first orthogonalizing $\tilde{\lambda}_i$ with respect to $\nabla_{\beta} \tilde{m}_i$. Let \tilde{E}_i be the TxQ matrix residuals from the matrix regression

$$\tilde{\lambda}_i \text{ on } \nabla_{\beta} \tilde{m}_i, \quad i=1, \dots, N. \quad (5.19)$$

Then compute $H = N - \text{SSR}$ from the regression

$$1 \text{ on } \tilde{u}_i' \tilde{E}_i, \quad i=1, \dots, N; \quad (5.20)$$

under (5.5), $H \stackrel{a}{\sim} \chi_Q^2$.

Because the moment assumptions (5.1)-(5.3) encompass HHG's FENB model, a rejection of (5.5) based on H necessarily implies misspecification of the FENB specification. A rejection implies some failure of (5.1)-(5.3), so one needs to work harder in specifying $E(y_i | x_i, \phi_i, \varphi_i)$ and $V(y_i | x_i, \phi_i, \varphi_i)$.

If the Hausman statistic fails to reject one might conclude that the first two moments in the latent variable model are correctly specified (this assumes that the Hausman test has power against interesting departures from (5.1)-(5.3)). If the QCMLE estimates are reasonably precise, then one could stop here. However, if (2.1) and (2.2) fail to hold, the QCMLEs could (but need not) have large standard errors.

Before searching for a more efficient estimator, it is useful to have direct evidence concerning the appropriateness of the multinomial distribution; if $E(y_i | n_i, x_i)$ and $V(y_i | n_i, x_i)$ match the first two moments of the Multinomial($n_i, p_1(x_i, \beta_0), \dots, p_T(x_i, \beta_0)$) distribution, worthwhile efficiency gains over the multinomial QCMLE are likely to be difficult to realize. A comparison of the usual and robust standard errors provides some guidance. A more formal test is HHG's serial correlation test for the FEP

model. However, the form of White's (1982) information matrix test covered by Wooldridge (1991) has some potential advantages. First, it imposes correctness of only the first two conditional moments (in this case $E(y_i|n_i, x_i)$ and $V(y_i|n_i, x_i)$) under H_0 , but it is asymptotically equivalent to nonrobust forms which take the entire distribution to be correctly specified. Second, it uses an estimate of the expected Hessian, \hat{A}_N , in its construction. Consequently, this test probably has better finite sample properties than HHG's outer product test; the latter is known to reject far too often in many situations.

The IM test for the multinomial model follows from Wooldridge (1991, Procedure 4.1), with a slight modification due to the singularity of $E(u_i u_i' | n_i, x_i)$. Let $\Omega(n_i, x_i, \beta)$ denote the $T \times T$ covariance matrix of the Multinomial($n_i, p_1(x_i, \beta), \dots, p_T(x_i, \beta)$) distribution. The t^{th} diagonal element of $\Omega(n_i, x_i, \beta)$ is $n_i p_t(x_i, \beta) [1 - p_t(x_i, \beta)]$ while the $(r, t)^{\text{th}}$ element is $-n_i p_r(x_i, \beta) p_t(x_i, \beta)$. Let $\hat{\Omega}_i = \Omega(n_i, x_i, \hat{\beta}_N)$ and

$$\hat{\xi}_i = (\text{vec}(\hat{u}_i \hat{u}_i' - \hat{\Omega}_i))' [\hat{V}_i \otimes \hat{V}_i]^{-1} \hat{\Lambda}_i - \hat{s}_i' \hat{A}_N^{-1} \hat{J}_N, \quad (5.21)$$

where

$$\hat{J}_N = N^{-1} \sum_{i=1}^N \nabla_{\beta} \hat{\Omega}_i' [\hat{V}_i \otimes \hat{V}_i]^{-1} \hat{\Lambda}_i$$

is $P \times Q$, $\nabla_{\beta} \hat{\Omega}_i(\beta) = \partial \text{vec}(\hat{\Omega}_i(\beta)) / \partial \beta$ is $T^2 \times P$, and $\hat{\Lambda}_i = \Lambda_i(x_i, \hat{\beta}_N)$ is a $T^2 \times Q$ matrix of selected linearly independent columns of the $T^2 \times P^2$ matrix $[\nabla_{\beta}^m \hat{\Omega}_i \otimes \nabla_{\beta}^m \hat{\Omega}_i]$. The IM test statistic is $IM = NR_u^2 = N - SSR$ from the regression

$$1 \text{ on } \hat{\xi}_i, \quad i=1, \dots, N. \quad (5.22)$$

Under the hypothesis that $E(y_i|n_i, x_i)$ and $V(y_i|n_i, x_i)$ match the first two moments of the Multinomial($n_i, p_1(x_i, \beta_0), \dots, p_T(x_i, \beta_0)$) distribution, $IM \xrightarrow{a} \chi_Q^2$.

If the IM test rejects, the search for more efficient estimates can proceed along two lines. First, one might estimate HHG's FENB model. But HHG's FENB model imposes $\varphi_i = 1 + \phi_i$ and, even if this restriction holds, the FENB QCMLE apparently does not enjoy the robustness properties of the FEP QCMLE. Because the quasi-score for the FENB cannot be expressed as in (3.13), the FENB QCMLE is generally inconsistent for β_0 unless (2.8) and (2.9) hold with $\alpha_i = 1$. For example, if the FEP model (2.1) and (2.2) holds, the FENB QCMLE is inconsistent for β_0 . Consequently, the FEP QCMLE is preferred to the FENB QCMLE.

A second approach is to construct a minimum chi-square estimator that is more efficient than the multinomial QCMLE but is nevertheless consistent under (5.1)-(5.3). There are a variety of minimum chi-square estimators that meet these criteria. Here I cover only one example, namely an estimator that combines the orthogonality conditions implied by QCMLE and MNLS. In the notation of section 3, Let $\hat{\phi}_N = \ddot{\beta}_N$ be a preliminary consistent estimator of β_0 (typically the multinomial QCMLE). Define $z_i = (1, n_i)'$,

$$L(x_i, \hat{\phi}_N) = \begin{bmatrix} W_i(\ddot{\beta}_N) \nabla_{\beta} p_i(\ddot{\beta}_N) & 0 \\ 0 & \nabla_{\beta} p_i(\ddot{\beta}_N) \end{bmatrix}, \quad (5.23)$$

which is $2T \times 2P$, and

$$\hat{\Xi}_N = \left[N^{-1} \sum_{i=1}^N L(x_i, \ddot{\beta}_N)' [z_i \otimes I_T] u_i(\ddot{\beta}_N) u_i(\ddot{\beta}_N)' [z_i' \otimes I_T] L(x_i, \ddot{\beta}_N) \right]^{-1}, \quad (5.24)$$

which is $2P \times 2P$. The residual function $u_i(\beta) = y_i - p_i(\beta)n_i$ can be expressed as $u_i(\beta) = y_i - C_i(\beta)z_i$, where $C_i(\beta) = [0 \mid p_i(\beta)]$ is $T \times 2$. Note that $\nabla_{\beta} C_i(\beta)' = [0 \mid \nabla_{\beta} p_i(\beta)']$, which is $P \times 2T$.

Denote the minimum chi-square estimator by $\hat{\beta}_N$. Then the asymptotic variance of $\hat{\beta}_N$ is consistently estimated by $(R_N' \hat{\Xi}_N R_N)^{-1}/N$, where

$$\hat{R}_N = N^{-1} \sum_{i=1}^N \hat{L}'_i [z_i z_i' \otimes I_T] \nabla_{\beta} \hat{C}_i.$$

The overidentification test statistic $Nr_N(\hat{\beta}_N)$ (see (4.6)) is asymptotically χ^2_P under (5.5), and provides additional evidence on the appropriateness of (5.1)-(5.3).

This section is concluded with a robust research methodology for count panel data models:

(i) Estimate model (5.1)-(5.3) by multinomial QCMLE. Compute the robust standard errors for $\hat{\beta}_N$.

(ii) Compute the robust Hausman specification test as in (5.19) and (5.20). If H rejects, conclude that (5.1)-(5.3) is misspecified.

(iii) If the Hausman test in (ii) fails to reject, compute the information matrix test as in (5.21) and (5.22). If the IM test fails to reject, conclude that the multinomial distribution adequately describes $E(y_i | n_i, x_i)$ and $V(y_i | n_i, x_i)$. The efficiency gains from minimum chi-square estimation are unlikely to be worthwhile.

(iv) If the IM test in (iii) rejects, a minimum chi-square procedure might produce noticeably tighter estimates. Compute the overidentification test statistic as further evidence on model specification.

6. Application to Gamma-type Unobserved Components Models

This section briefly outlines how the CLP approach can be applied to models where the conditional variance is proportional to the square of the conditional mean. Some popular continuous distributions for nonnegative variables, in particular the lognormal and the gamma, have first two moments corresponding to this assumption. For $t=1, \dots, T$, assume that

$$E(y_{it}|x_i, \phi_i) = \phi_i \mu(x_{it}, \beta_0) \quad (6.1)$$

$$V(y_{it}|x_i, \phi_i) = \sigma_0^2 [E(y_{it}|x_i, \phi_i)]^2 = \sigma_0^2 [\phi_i \mu(x_{it}, \beta_0)]^2 \quad (6.2)$$

$$CV(y_{it}, y_{ir}|x_i, \phi_i) = 0, \quad t \neq r. \quad (6.3)$$

Assumption (6.1) is essentially the same as (5.1), while (6.3) corresponds to (5.3). Note that only one unobserved effect, ϕ_i , is allowed. I do not know how to allow the proportionality parameter σ_0^2 to vary across i . Equation (6.2) corresponds to what statisticians refer 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:

$$SD(y_{it}|x_i, \phi_i)/E(y_{it}|x_i, \phi_i) = \sigma_0. \quad (6.4)$$

As far as I know, there has been no work analyzing such models in an unobserved components, panel data setting, with or without distributional assumptions. This is probably because, when y_{it} is a nonnegative continuous random variable, ~~most~~ researchers use $\log(y_{it})$ in a linear fixed effects model. But if interest lies in $E(y_{it}|x_i, \phi_i)$, (6.1)-(6.3) might be preferred; additional assumptions about $D(y_{it}|x_i, \phi_i)$ are needed to recover $E(y_{it}|x_i, \phi_i)$ from $E[\log(y_{it})|x_i, \phi_i]$.

The conditioning that eliminates the individual effect ϕ_i is more restrictive than that used in section 5. Let $n_i = \sum_{t=1}^T y_{it}$ be as before. Then, for each t , the linear predictor of y_{it} on n_i , conditional on (x_i, ϕ_i) , is:

$$L(y_{it}|n_i; x_i, \phi_i) = \frac{E(y_{it}n_i|x_i, \phi_i)}{E(n_i^2|x_i, \phi_i)} n_i. \quad (6.5)$$

But

$$E(y_{it}n_i|x_i, \phi_i) = CV(y_{it}, n_i|x_i, \phi_i) + E(y_{it}|x_i, \phi_i)E(n_i|x_i, \phi_i)$$

$$\begin{aligned}
& - V(y_{it}|x_i, \phi_i) + [\phi_i \mu(x_{it}, \beta_o)] \left[\sum_{r=1}^T \phi_i \mu(x_{ir}, \beta_o) \right] \\
& - \phi_i^2 \left[\sigma_o^2 [\mu(x_{it}, \beta_o)]^2 + \mu(x_{it}, \beta_o) \left(\sum_{r=1}^T \mu(x_{ir}, \beta_o) \right) \right].
\end{aligned}$$

Similarly,

$$\begin{aligned}
E(n_i^2|x_i, \phi_i) &= V(n_i|x_i, \phi_i) + [E(n_i|x_i, \phi_i)]^2 \\
&= \phi_i^2 \left[\sigma_o^2 \sum_{r=1}^T [\mu(x_{ir}, \beta_o)]^2 + \left(\sum_{r=1}^T \mu(x_{ir}, \beta_o) \right)^2 \right].
\end{aligned}$$

Therefore,

$$L(y_{it}|n_i; x_i, \phi_i) \equiv p_t(x_i, \theta_o) n_i, \quad (6.6)$$

where $\theta_o \equiv (\beta_o', \sigma_o^2)'$ and

$$p_t(x_i, \theta_o) \equiv \frac{\sigma_o^2 [\mu(x_{it}, \beta_o)]^2 + \mu(x_{it}, \beta_o) \left(\sum_{r=1}^T \mu(x_{ir}, \beta_o) \right)}{\sigma_o^2 \sum_{r=1}^T [\mu(x_{ir}, \beta_o)]^2 + \left(\sum_{r=1}^T \mu(x_{ir}, \beta_o) \right)^2}. \quad (6.7)$$

Note that

$$\sum_{t=1}^T p_t(x_i, \theta) = 1 \text{ for all } \theta.$$

In terms of the vector y_i , (6.6) is expressed as

$$L(y_i|n_i; x_i, \phi_i) = p(x_i, \theta_o) n_i, \quad (6.8)$$

where $p(x_i, \theta) \equiv (p_1(x_i, \theta), \dots, p_T(x_i, \theta))'$.

Although the right hand side of (6.8) is of the same form as (5.5), there is an important difference. Under (6.1)-(6.3), (6.8) does not also represent $L(y_i|1, n_i; x_i, \phi_i)$. In fact, the CLP of y_i on unity and n_i depends on ϕ_i , rendering it useless for estimating σ_o^2 and β_o . Taking $z_i \equiv n_i$ in section 3 restricts the class of consistent estimators for σ_o^2 and β_o . Nevertheless, there are plenty of orthogonality conditions of the form

$$E[D(x_i, \theta_o, \gamma^*)' n_i (y_i - p(x_i, \theta_o) n_i)] = 0 \quad (6.9)$$

to identify σ_o^2 and β_o .

Weighted MNLS estimators, which solve

$$\min_{\theta \in \Theta} \sum_{i=1}^N (y_i - p(x_i, \theta) n_i)' G(x_i, \hat{\gamma}_N) (y_i - p(x_i, \theta) n_i), \quad (6.10)$$

are generally consistent and asymptotically normally distributed. Further, given one such estimator, it is straightforward to stack WMNLS orthogonality conditions to obtain a more efficient minimum chi-square estimator. The robust, regression-based Hausman test for comparing two WMNLS estimators covered by Wooldridge (1991) is a special case of the tests in section 4, and can be used to test the validity of (6.8).

7. Models with Serial Correlation

For some applications, the zero covariance assumptions (2.12) and (6.3) might be too restrictive (although recall that these are conditional on latent effects). For the model in section 6, it is straightforward to relax the zero covariance assumption. In fact, (6.2) and (6.3) can be replaced with the general assumption

$$V(y_i | x_i, \phi_i) = \phi_i^2 \Omega(x_i, \delta_o), \quad (7.1)$$

where $\Omega(x_i, \delta)$ is a $T \times T$ positive definite variance function. The conditioning argument used in section 6 still eliminates ϕ_i . In fact, letting $\Omega_t(x_i, \delta)$ denote the t^{th} row of $\Omega(x_i, \delta)$ and $j_T = (1, 1, \dots, 1)'$, $L(y_{it} | n_i; x_i, \phi_i)$ is given by (6.6) with

$$p_t(x_i, \theta_o) = \frac{\Omega_t(x_i, \delta_o) j_T + \mu(x_{it}, \beta_o) \left(\sum_{r=1}^T \mu(x_{ir}, \beta_o) \right)}{j_T' \Omega(x_i, \delta_o) j_T + \left(\sum_{r=1}^T \mu(x_{ir}, \beta_o) \right)^2}. \quad (7.2)$$

Model (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

$$\Omega(x_i, \delta_o) = \sigma_o^2 \Delta(x_i, \beta_o) R_T(\rho_o) \Delta(x_i, \beta_o), \quad (7.3)$$

where

$$\Delta(x_i, \beta_o) = \text{diag}\{\mu(x_{i1}, \beta_o), \dots, \mu(x_{iT}, \beta_o)\} \quad (7.4)$$

and $R_T(\rho)$ is the $T \times T$ matrix with $(r, t)^{\text{th}}$ element $\rho^{|r-t|}$. This model maintains (6.2) but relaxes (6.3) to $CV(y_{it}, y_{ir} | x_i, \phi_i) = \rho_o \mu(x_{it}, \beta_o) \mu(x_{ir}, \beta_o)$.

Model (6.1) and (7.1) can also allow for serial correlation in count-type models, but the individual dispersion is restricted in this case. A model with constant AR(1) serial correlation chooses $\Omega(x_i, \delta_o)$ as in (7.3), except that

$$\Delta(x_i, \beta_o) = \text{diag}\{[\mu(x_{i1}, \beta_o)]^{1/2}, \dots, [\mu(x_{iT}, \beta_o)]^{1/2}\}. \quad (7.5)$$

In terms of model (2.10)-(2.12), (2.11) has been maintained and (2.12) has been relaxed at the cost of imposing $\varphi_i = \sigma_o^2 \phi_i$ (compare to the HHG FENB assumption $\varphi_i = 1 + \phi_i$).

The identification issue in these more complicated models warrants some attention. As was seen in section 2, σ_o^2 is not identified if $\rho_o = 0$, in which case φ_i is free to vary independently of ϕ_i . Also, when $\rho_o \neq 0$, the CLP $L(y_i | n_i; x_i, \phi_i)$ must be used to estimate β_o , ρ_o , and σ_o^2 ; the CLP

$L(y_i | 1, n_i; x_i, \phi_i)$ now depends on ϕ_i and is therefore useless. Thus, the multinomial QCMLE, which is consistent under (2.10)-(2.12), is no longer consistent when $\rho_0 \neq 0$. A weighted least squares procedure or GMM estimator must be used instead with $p_t(x_i, \theta_0)$ given by (7.2).

8. Concluding Remarks

This paper has shown how the notion of a conditional linear predictor can be used to eliminate individual components in certain classes of multiplicative unobserved effects models. This technique can be viewed as a particular implementation of the general approach suggested by Neyman and Scott (1948) for obtaining consistent estimates of fixed dimensional parameters in the presence of an infinite dimensional nuisance parameter.

The first two moments of the count model in section 5 should be general enough for many applications. A corollary of the analysis is that the multinomial QCMLE has important robustness properties, and can be used to consistently estimate the parameters of a fairly general mean and variance function. Nevertheless, obtaining minimum chi-square estimates could result in efficiency gains. The model of section 6, intended for continuous, nonnegative variables, can be used in place of the usual practice of taking natural logs and postulating a linear model.

The models in sections 5, 6, and 7 assume that x_i is strictly exogenous conditional on the latent variable or variables. This rules out feedback from y_{it} to x_{ir} , $r > t$ (i.e., $\{y_{it}\}$ cannot Granger-cause $\{x_{it}\}$). While this 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 so,

all of the estimators considered in section 5 are generally inconsistent. Future research could usefully investigate how to relax the strict exogeneity assumption in nonlinear unobserved components models.

Finally, conditional linear predictors can also be used to robustly estimate multiplicative unobserved components models for multivariate time series. For example, suppose that $((y_t, x_t): t=1, 2, \dots)$ is a vector time series, with y_t a $J \times 1$ vector of counts and x_t a $K \times 1$ vector of conditioning variables. A multiplicative unobserved components model might specify an analog of (2.1) and (2.2): for $j=1, \dots, J$,

$$y_{tj} | x_t, \phi_t \sim \text{Poisson}(\phi_t \mu_j(x_t, \beta_0)) \quad (8.1)$$

$$y_{tj}, y_{th} \text{ are independent, conditional on } x_t, \phi_t, j \neq h. \quad (8.2)$$

Any dependence between y_{tj} and y_{th} is due entirely to the unobserved (or "common") component ϕ_t . If (8.1) and (8.2) hold, the conditioning argument used in section 2 can be used to eliminate ϕ_t . Then β_0 can be estimated by CMLE (although the score of the log-likelihood would not necessarily be a martingale difference sequence). Alternatively, the moment assumptions

$$E(y_{tj} | x_t, \phi_t, \varphi_t) = \phi_t \mu_j(x_t, \beta_0) \quad (8.3)$$

$$V(y_{tj} | x_t, \phi_t, \varphi_t) = \varphi_t \phi_t \mu_j(x_t, \beta_0) \quad (8.4)$$

$$CV(y_{tj}, y_{th} | x_t, \phi_t, \varphi_t) = 0, j \neq h \quad (8.5)$$

can be used. If $n_t \equiv \sum_{j=1}^J y_{tj}$, then the linear predictor of y_t on $(1, n_t)'$, conditional on (x_t, ϕ_t, φ_t) , eliminates the unobservables ϕ_t and φ_t , as before. Generally, unless (8.3)-(8.5) represent completely specified dynamics, the theory of section 3 must be extended to allow the score to be serially correlated. But this is no more difficult than QCMLE or GMM with incompletely specified dynamics.

1. The term "conditional maximum likelihood" is somewhat unfortunate in light of modern econometrics. This is because virtually all maximum likelihood is conditional on a set of explanatory variables. The phrase "conditional maximum likelihood" appeared in the statistics literature where explanatory variables are treated as nonrandom. Thus, one specifies an unconditional joint distribution for the explained variables (which depends on the explanatory variables), and then conditions on a function of the explained variables. Because it is too late to change the terminology, I stick to standard usage; the term conditional MLE is reserved for the case when a function of the explained variables is conditioned on.

2. For linear models, Chamberlain (1982) uses a linear projection argument, which imposes no restrictions on the distribution of ϕ_i given x_i . Due to the nonlinear nature of the current models, this approach is unavailable.

3. If the asymptotic variance of $\sqrt{N}(\hat{\theta}_N - \theta_0)$ is V_0 , it is natural to define V_0/N to be the asymptotic variance of $\hat{\theta}_N$, denoted $AV(\hat{\theta}_N)$. If \hat{V}_N is a consistent estimator of V_0 , then \hat{V}_N/N is said to be an estimator of $AV(\hat{\theta}_N)$.

4. HHG introduce two unobserved effects, which they label μ_i and ϕ_i . However, it is easily seen from their equations (HHG (1984, p. 924)) that their FENB model is equivalent to (2.8)-(2.9) with $\alpha_i = 1$.

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: Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics*, Volume I, Amsterdam: North Holland.
- Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics* 34, 305-334.
- Gourieroux, C., A. Monfort, and C. Trognon (1984), "Pseudo-Maximum Likelihood Methods: Theory," *Econometrica* 52, 681-700.
- Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029-1054.
- Hausman, J.A. (1978), "Specification Tests in Econometrics," *Econometrica* 46, 1251-1271.
- Hausman, J.A., B.H. Hall, and Z. Griliches (1984), "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica* 52, 909-938.
- McCullagh, P., and J.A. Nelder (1989), *Generalized Linear Models*, New York: Chapman and Hall.
- Newey, W.K. (1985), "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica* 53, 1047-1070.
- Newey, W.K. (1987), "Efficient Estimation of Semiparametric Models via Moment Restrictions," manuscript, Princeton University.
- Neyman, J. and E.L. Scott (1948), "Consistent Estimates Based on Partially Consistent Observations," *Econometrica* 16, 1-32.
- 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. (1989), "Interstate Business Tax Differentials and New Firm Location: Evidence from Panel Data," manuscript, MIT Sloan School of Management.

- White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* 50, 1-26.
- Wooldridge, J.M. (1990), "A Unified Approach to Robust, Regression-Based Specification Tests," *Econometric Theory* 6, 17-43.
- Wooldridge, J.M. (1991), "Specification Testing and Quasi-Maximum Likelihood Estimation," forthcoming, *Journal of Econometrics* 47.



Date Due

MAY 11 1961

APR. 04 1961

JUL 30 1961

MIT LIBRARIES DUPL



3 9080 00701592 5

