cemmap

centre for microdata methods and practice

Feedback in panel data models

Gary Chamberlain

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP34/21

Economic and Social Research Council

Department of Economics
Harvard University

September 1993

FEEDBACK IN PANEL DATA MODELS

Gary Chamberlain*

Correspondence to: Gary Chamberlain
Department of Economics
Harvard University
Cambridge, MA 02138
USA

ABSTRACT

Much of the analysis of panel data has been based on an assumption of strict exogeneity. Distributions are specified for outcome variables conditional on a latent individual effect and conditional on observed predictor variables at all dates, with the future values of the predictor variables assumed to have no effect on the conditional distribution. The paper relaxes this assumption in order to allow for lagged dependent variables and, more generally, for feedback from lagged dependent variables to current values of the predictor variables. Such feedback would arise in an evaluation study if the treatment variable is randomly assigned only conditional on the individual effect and on previous outcomes.

An information bound is derived for a semiparametric regression model with sequential moment restrictions, with the information set increasing over time. The bound is then applied to a model with a (scalar) multiplicative random effect. The mean of the random effect conditional on the predictor variables is not restricted, so that the random effect can control for various omitted variables. This conditional mean is the nonparametric component of the semiparametric regression model. There is a transformation that eliminates the random effect and leads to a set of sequential moment restrictions in which the moment function depends on only a finite-dimensional parameter. The information bound for this simpler problem coincides with that of the original problem. The form of the optimal instrumental variables is derived.

The paper also considers the identification problems that arise when the random effect is a vector with two or more components.

FEEDBACK IN PANEL DATA MODELS

1. INTRODUCTION

Much of the analysis of panel data models has been based on an assumption of strict exogeneity. A linear regression example is

$$E(y_{it} \,|\, x_{i1}, \ldots, x_{iT}, c_i) = \beta x_{it} + c_i \qquad (i = 1, \ldots, n;\; t = 1, \ldots, T).^1$$

We observe $(x_{i1}, y_{i1}, \ldots, x_{iT}, y_{iT})$ for a random sample of n individuals. The individual effect c_i is not observed. The predictor variable x is strictly exogenous in that its values at all dates t are simultaneously conditioned on in the regression specification. It then follows that

$$E(y_{it} - y_{i,t+1} \,|\, x_{i1}, \ldots, x_{iT}) = \beta(x_{it} - x_{i,t+1}).$$

So we have a regression specification that only involves observables, and a least-squares regression of, for example, $y_{i1} - y_{i2}$ on $x_{i1} - x_{i2}$ provides a consistent (as $n \to \infty$) estimate of β.

In binary response models, strict exogeneity plays a key role in Rasch's (1960, 1961) conditional likelihood approach to logistic models. Strict exogeneity is also necessary in Manski's (1987) application of the maximum score estimator to models with unobserved individual effects. Hausman, Hall, and Griliches (1984) consider panel data models for a count variable. Their conditional likelihood estimators require strict exogeneity. The need for distributional assumptions in those models is relaxed in Wooldridge (1990), but strict exogeneity is still required. Likewise, strict exogeneity is needed in Chamberlain's (1992a) analysis of random coefficient models. Finally, Honoré's (1992) treatment of panel data models with censoring relies upon strict exogeneity.[2]

We would like to relax the strict exogeneity assumption and allow for lagged dependent variables and, more generally, for feedback from lagged dependent variables to current values of the predictor variables. A linear regression example is

$$E(y_{it} \,|\, x_{i1}, y_{i0}, \ldots, x_{it}, y_{i,t-1}, c_i) = \beta x_{it} + \gamma y_{i,t-1} + c_i. \tag{1.1}$$

This specification allows for x_t to depend upon lagged y, but the feedback is not given a parametric form.

A similar specification could arise in an evaluation study, where y_{it} is earnings and x_{it} indicates whether or not the individual has received job training as of date t. Let y_{it}^0 denote earnings in the absence of training, and suppose that the training is administered in period s, with $x_{it} = 0$ for $t < s$ and $x_{it} = x_{is}$ for $t \geq s$. If the training effect is the same for everyone, we have

$$y_{it} = y_{it}^0 + \beta_t x_{it} \qquad (t = 1, \ldots, s-1, s+1, \ldots, T).$$

It is commonly observed that there is a dip in the pretraining earnings of participants in job training programs (Ashenfelter (1978), Ashenfelter and Card (1985), LaLonde (1986)). So we allow selection into training to depend upon previous earnings, as well as upon an individual effect c_i. Training is randomly assigned conditional upon previous earnings and c_i, in the following sense:

$$E(y_{it}^0 \,|\, y_{i0}^0, \ldots, y_{i,t-1}^0, c_i, x_{it}) = E(y_{it}^0 \,|\, y_{i0}^0, \ldots, y_{i,t-1}^0, c_i).$$

To complete the model, we use the following specification for the dependence of y_{it}^0 on lagged values and c_i:

$$E(y_{it}^0 \,|\, y_{i0}^0, \ldots, y_{i,t-1}^0, c_i) = \gamma y_{i,t-1}^0 + c_i.$$

Then we have

$$E(y_{it} \,|\, y_{i0}, \ldots, y_{i,t-1}, c_i, x_{it}) = \gamma y_{i,t-1} + c_i + (\beta_t - \gamma \beta_{t-1}) x_{it} \tag{1.2}$$

$(t = 1, \ldots, s-1, s+2, \ldots, T)$,

$$E(y_{i,s+1} \,|\, y_{i0}, \ldots, y_{i,s-1}, c_i, x_{i,s+1}) = \gamma^2 y_{i,s-1} + (1+\gamma) c_i + \beta_{s+1} x_{i,s+1}.$$

Models with lagged dependent variables and individual effects have been considered by, among others, Balestra and Nerlove (1966), Anderson and Hsiao (1982), Bhargava and Sargan (1983), Chamberlain (1984), Holtz-Eakin, Newey, and Rosen (1988), Arellano and Bond (1991), Arellano and Bover (1993), and Ahn and Schmidt (1993). Keane and Runkle (1992) consider relaxing strict exogeneity in a panel data model with an additive individual effect; their motivation is work on the permanent income hypothesis and liquidity constraints, as in Zeldes (1989).

We shall work with the following extension of (1.1):

$$d_t(z_i, \theta_0) = R_t(w_i^t, \theta_0)c_i + v_{it} \tag{1.3}$$

$$E(v_{it} \mid w_i^t) = 0 \qquad (i = 1, \ldots, n;\; t = 1, \ldots, T).$$

Here z_i contains the data on the i^{th} unit; $w_i^t \equiv (w_{i1}, \ldots, w_{it})$ is contained in z_i. The scalar random effect c_i is not observed. d_t and R_t are given functions. We are interested in θ_0 and in $\phi_0 \equiv E(c_i)$. The relationship between the random effect and the predictor variables is not restricted; in particular, we do not assume that $E(c_i \mid w_i^T) = E(c_i)$. In this sense, the "fixed effects" terminology would be appropriate. The feature that makes this nonlinear model tractable is that the scalar random effect multiplies a known function of the data w_i^t and the parameter θ_0.

One example is a linear model in which interactions are allowed:

$$y_{it} = \theta_{01}x_{it} + \theta_{02}y_{i,t-1} + c_i + \theta_{03}x_{it}c_i + v_{it}.$$

Here we might have $w_{it} = (x_{it}, y_{i,t-1})$ (with y_{i0} observed); we set $d_t(z_i, \theta) = y_{it} - \theta_1 x_{it} - \theta_2 y_{i,t-1}$, and $R_t(w_i^t, \theta) = 1 + \theta_3 x_{it}$.

Another example, which would be of interest when y is a nonnegative random variable such as a count, is

$$y_{it} = \exp(\theta_{01}x_{it} + \theta_{02}y_{i,t-1} + \alpha_i) + v_{it},$$

with $w_{it} = (x_{it}, y_{i,t-1})$. Here we set $d_t(z_i, \theta) = y_{it}$, $R_t(w_i^t, \theta) = \exp(\theta_1 x_{it} + \theta_2 y_{i,t-1})$, and $c_i = \exp(\alpha_i)$.

Our treatment of the random effects model in (1.3) is based on a semiparametric regression model with sequential moment restrictions. That model, which is set up in Section 2, has the following form:

$$E[\rho_t(z_i, \theta_0, h_0(w_i^T)) \,|\, w_i^t] = 0 \qquad (t = 1, \ldots, T)$$

for some $\theta_0 \in \mathcal{R}^p$ and some scalar-valued function h_0. Here ρ_t is a given function with domain a subset of $\mathcal{R}^m \times \mathcal{R}^p \times \mathcal{R}$. We derive an efficiency bound for θ_0 and for $\phi_0 \equiv E[h_0(w_i^T)]$. These results are then applied in Section 3 to the multiplicative random effects model. The link between the models is provided by setting $h_0(w_i^T) = E(c_i \,|\, w_i^T)$.

Section 4 considers models in which the random effect is a vector with two or more components. An example is

$$\begin{aligned} y_{it} &= \theta_0 y_{i,t-1} + c_{1i} + c_{2i} x_{it} + v_{it} \\ E(v_{it} \,|\, x_{i1}, y_{i0}, \ldots, x_{it}, y_{i,t-1}) &= 0. \end{aligned} \tag{1.4}$$

We show in Section 4 that although there can be positive results for some special cases, in general there are severe identification problems in such models.

2. SEQUENTIAL MOMENT RESTRICTIONS

The observations on the i^{th} unit are contained in a $m \times 1$ vector z_i. We assume that $\{z_i\}_{i=1}^{\infty}$ is a sequence of i.i.d. random vectors with probability measure P. The random vectors $w_{i1}, \ldots, w_{iT}$ are components of z_i. We shall simplify the notation by letting $z, w_1, \ldots, w_T$ denote random vectors whose joint distribution coincides with that of $z_i, w_{i1}, \ldots, w_{iT}$.

We are given functions $\rho_t \colon Z \times \Theta \times N \to \mathcal{R}$, where $Z \subset \mathcal{R}^m$ contains the support of P, the parameter space Θ is an open subset of $\mathcal{R}^p$, and N is an open subset of $\mathcal{R}$. We assume that $\rho_t(\cdot, \theta, \cdot)$ is a measurable function for each $\theta \in \Theta$, and that $\rho_t(a, \cdot, \cdot)$ is continuously differentiable on $\Theta \times N$ for each $a \in Z$. We shall consider triples (Q, θ, h) that satisfy the following condition:

CONDITION (C). (i) Q is a probability measure whose support is a subset of Z; $\theta \in \Theta$; h is a measurable function mapping the Q-support of $(w_1, \ldots, w_T)$ into N. (ii) For $t = 1, \ldots, T$,

$$E_Q[\rho_t(z, \theta, h(w^T)) \,|\, w^t] = 0,$$

where $w^t \equiv (w_1, \ldots, w_t)$.

Suppose that (Q, θ, h) satisfies Condition (C). The analysis is simplified by the following transformation:

$$g_t(z, \alpha, \beta) \equiv \rho_t(z, \alpha, \beta) - H_t H_T^{-1} \rho_T(z, \alpha, \beta) \qquad (t = 1, \ldots, T-1),$$

where

$$H_t = E_Q[\partial \rho_t(z, \theta, h(w^T))/\partial h \,|\, w^T] \quad (t = 1, \ldots, T)$$

and we assume that $Q\{H_T = 0\} = 0$. This transformation is useful because

$$E_Q[\partial g_t(z, \theta, h(w^T))/\partial h \,|\, w^T] = 0. \tag{2.1}$$

After making the transformation, we apply a forward filter:

$$\tilde{g}_{T-1}(z,\alpha,\beta) = g_{T-1}(z,\alpha,\beta) \tag{2.2}$$

$$\tilde{g}_s(z,\alpha,\beta) = g_s(z,\alpha,\beta) - \Gamma_{s,s+1}\tilde{g}_{s+1}(z,\alpha,\beta) - \ldots - \Gamma_{s,T-1}\tilde{g}_{T-1}(z,\alpha,\beta) \qquad (s = T-2,\ldots,1),$$

where

$$\Gamma_{st} = E_Q(g_s\tilde{g}_t \mid w^t)[E_Q(\tilde{g}_t^2 \mid w^t)]^{-1} \qquad (s < t)$$

with g_s and $\tilde{g}_t$ evaluated at $(z,\theta,h(w^T))$.

The idea of the forward filter comes from Hayashi and Sims (1983). In our case the coefficients Γ_{st} in the forward filter are stochastic, as in Hansen, Heaton, and Ogaki (1988). The advantage of the forward filter is that it results in moment functions that are conditionally orthogonal given the larger (of s and t) information set:

$$E_Q[\tilde{g}_s(z,\theta,h(w^T))\tilde{g}_t(z,\theta,h(w^T)) \mid w^t] = 0 \qquad (s < t). \tag{2.3}$$

The forward filter can be shown to solve the following prediction problem:

$$\min_{A_{s+1},\ldots,A_{T-1}} E_Q(g_s - \hat{g}_s)^2$$

subject to

$$\hat{g}_s = A_{s+1}(w^{s+1})g_{s+1} + \ldots + A_{T-1}(w^{T-1})g_{T-1}$$

is solved by setting

$$\hat{g}_s = \Gamma_{s,s+1}\tilde{g}_{s+1} + \Gamma_{s,s+2}\tilde{g}_{s+2} + \ldots + \Gamma_{s,T-1}\tilde{g}_{T-1}$$

(where the g's and $\tilde{g}$'s are evaluated at $(z,\theta,h(w^T))$).

Define $\phi = E_Q[h(w^T)]$. The following condition on the triple (Q,θ,h) will insure that the information matrix for (θ,ϕ) is positive definite.

CONDITION (PD). (i) $E_Q[\rho_t^2(z,\theta,h(w^T))] < \infty$ $(t = 1,\ldots,T)$ and $E_Q[h^2(w^T)] < \infty$. (ii) Let $q_0 = h(w^T) - \phi$ and let $q_t = \rho_t(z,\theta,h(w^T))$ for $t = 1,\ldots,T$. Let $\tilde{q}_T = q_T$ and

$$\tilde{q}_s = q_s - \Psi_{s,s+1}\tilde{q}_{s+1} - \ldots - \Psi_{sT}\tilde{q}_T \qquad (s = T-1,\ldots,0),$$

where

$$\Psi_{st} = E_Q(q_s\tilde{q}_t \,|\, w^t)[E_Q(\tilde{q}_t^2 \,|\, w^t)]^{-1} \qquad (s < t).$$

Then $E_Q(\tilde{q}_t^2 \,|\, w^t) > 0$ with Q-probability one ($t = 0, \ldots, T$), where $w^0 \equiv 1$. (iii) H_T is nonzero with Q-probability one. (iv) Let

$$V(Q,\theta,h) = \begin{pmatrix} V_\theta & V_{\theta\phi} \\ V_{\phi\theta} & V_\phi \end{pmatrix} \tag{2.4}$$

$$V_\theta = \left[\sum_{t=1}^{T-1} E_Q(\tilde{G}_t' \tilde{\Sigma}_t^{-1} \tilde{G}_t)\right]^{-1}$$

$$V_{\phi\theta} = V_{\theta\phi}' = -KV_\theta$$

$$V_\phi = \mathrm{Var}_Q\left[h - H_T^{-1}\rho_T - \sum_{t=1}^{T-1} \Lambda_t \tilde{g}_t\right] + KV_\theta K',$$

where

$$\tilde{G}_t = E_Q(\partial\tilde{g}_t/\partial\theta' \,|\, w^t), \qquad \tilde{\Sigma}_t = E_Q(\tilde{g}_t^2 \,|\, w^t),$$

$$\Lambda_t = E_Q[(h - H_T^{-1}\rho_T)\tilde{g}_t \,|\, w^t]\tilde{\Sigma}_t^{-1}, \qquad D_T = E_Q(\partial\rho_T/\partial\theta' \,|\, w^T),$$

$$K = \sum_{t=1}^{T-1} E_Q(\Lambda_t \tilde{G}_t) + E_Q(H_T^{-1} D_T)$$

with $\tilde{g}_t$ and ρ_T evaluated at $(z, \theta, h(w^T))$. Then $V(Q,\theta,h)$ is well-defined; i.e., the relevant expectations exist and are finite, $Q\{\tilde{\Sigma}_t > 0\} = 1$, and $\sum_{t=1}^{T-1} E_Q(\tilde{G}_t' \tilde{\Sigma}_t^{-1} \tilde{G}_t)$ is positive definite.

We shall assume that (P, θ_0, h_0) satisfies Conditions (C) and (PD), and we let $\phi_0 = E_P[h_0(w^T)]$. Define

$$V_0 = V(P, \theta_0, h_0). \tag{2.5}$$

If the distribution of z is multinomial with known, finite support, then the estimation problem becomes parametric. The unknown parameters are the probabilities of the different values for z, and these probabilities are restricted by Condition (C). The following theorem is based on evaluating the Fisher information matrix in the multinomial case.

It shows that V_0 is the variance bound for (θ_0, ϕ_0). Theorem 2 in the Appendix uses multinomial approximation to extend the result to more general distributions.

THEOREM 1. Suppose that P has known, finite support and that (P, θ_0, h_0) satisfies Conditions (C) and (PD). Then evaluating the Fisher information matrix gives V_0 as the variance bound for (θ_0, ϕ_0). (Proof in the Appendix.)

To interpret the bound for θ_0, note that[3]

$$E[\tilde{g}_t(z, \theta_0, h_0(w^T)) \,|\, w^t] = 0 \qquad (t = 1, \ldots, T-1). \tag{2.6}$$

Since, from (2.1), $\tilde{g}_t$ does not depend upon h to first order, the information bound based on (2.6) in period t is $E(\tilde{G}_t' \tilde{\Sigma}_t^{-1} \tilde{G}_t)$—i.e., we proceed as if h_0 were known. Then, given the conditional orthogonality in (2.3), we simply add up the information bounds for each period.

In order to interpret the bound for ϕ_0, suppose that θ_0 is known—the term $K V_\theta K'$ accounts for the additional variance due to the estimation of θ_0. Then consider

$$\mathrm{Var}(h_0 - H_T^{-1} \rho_T) = \mathrm{Var}(h_0) + \mathrm{Var}(H_T^{-1} \rho_T)$$

(since $E(\rho_T \,|\, w^T) = 0$). The term $\mathrm{Var}(h_0)$ reflects the variance in estimating $E[h_0(w^T)]$ when h_0 is known and the w^T distribution is unknown. The term $\mathrm{Var}(H_T^{-1} \rho_T)$ accounts for the sampling variability in estimating h_0 when the w^T distribution is known, and when the estimation is based solely on the terminal period moment condition, $E(\rho_T \,|\, w^T) = 0$. To see this, consider the discrete case with mass points for w^T at $\tau_1, \ldots, \tau_J$. Then

$$E[\rho_T(z, \theta_0, h_0(\tau_j)) \,|\, w^T = \tau_j] = 0$$

suggests the sample analog

$$\frac{1}{n_j} \sum_{i: w_i^T = \tau_j} \rho_T(z_i, \theta_0, \hat{h}_0(\tau_j)) = 0$$

(with n_j equal to the number of observations with $w_i^T = \tau_j$), which leads to

$$\sqrt{n}[\hat{h}_0(\tau_j) - h_0(\tau_j)]$$
$$\cong -\left[\frac{1}{n_j}\sum_{i:w_i^T=\tau_j} \partial\rho_T(z_i,\theta_0,h_0(\tau_j))/\partial h\right]^{-1} \frac{\sqrt{n}}{n_j}\sum_{i:w_i^T=\tau_j} \rho_T(z_i,\theta_0,h_0(\tau_j)).$$

Then

$$\sqrt{n}\sum_{j=1}^{J} P\{w_i^T = \tau_j\}[\hat{h}_0(\tau_j) - h_0(\tau_j)]$$
$$\cong -\frac{1}{\sqrt{n}}\sum_{j=1}^{J} P\{w_i^T = \tau_j\}\frac{n}{n_j}\sum_{i:w_i^T=\tau_j} H_T^{-1}\rho_T(z_i,\theta_0,h_0(\tau_j))$$
$$\cong -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} H_T^{-1}\rho_T(z_i,\theta_0,h_0(w_i^T)).$$

Now note that $E(\tilde{g}_t \,|\, w^t) = 0$ implies that $E(\Lambda_t\tilde{g}_t) = 0$, provided that Λ_t is a function of w^t. So including the term $\sum_t \Lambda_t\tilde{g}_t$ does not add bias, and the weights Λ_t can be chosen to reduce variance. In fact we have

$$V_\phi = \min_{A_1,\ldots,A_{T-1}} E[h_0 - H_T^{-1}\rho_T - \phi_0 - \sum_{t=1}^{T-1} A_t(w^t)g_t]^2 + KV_\theta K'. \tag{2.7}$$

Here the variance reducing term is expressed in terms of the g_t's instead of the forward-filtered residuals $\tilde{g}_t$. The conditional orthogonality of the $\tilde{g}_t$ allows a simple explicit expression for the minimization problem in (2.7). So we can think of the terminal period T as providing a direct estimate of ϕ_0, say $\hat{\phi}_{(T)}$. Then the earlier periods are used to form mean-zero statistics that are correlated with $\hat{\phi}_{(T)}$; a linear combination of these statistics can be added to $\hat{\phi}_{(T)}$ to obtain the efficient estimate—see (3.3) in the next section for an example of this.

The variance bound in Theorem 1 extends the bound in Chamberlain (1992a) to the case of sequential moment restrictions. The framework here is, however, quite restrictive:

h is a scalar-valued function of w^T, instead of a vector whose components are allowed to depend on various functions of w^T. Nevertheless, this framework does encompass an interesting set of panel data models with multiplicative random effects. The next section shows how these panel data models fit into our semiparametric regression framework.

3. MULTIPLICATIVE RANDOM EFFECTS

The general model is

$$d_t(z_i,\theta_0) = R_t(w_i^t,\theta_0)c_i + v_{it} \tag{3.1}$$

$$E(v_{it}\,|\,w_i^t) = 0 \qquad (i=1,\ldots,n;\ t=1,\ldots,T),$$

where d_t and R_t are given functions, and c_i is a scalar latent variable that is free to vary over the cross section but is constant over time. We are interested in θ_0 and in $\phi_0 \equiv E(c_i)$.

This model fits into the Section 2 framework as follows: dropping the i subscripts, let

$$\rho_t(z,\theta,h) \equiv d_t(z,\theta) - R_t(w^t,\theta)h$$

$$h_0(w^T) \equiv E(c\,|\,w^T).$$

Then

$$\rho_t(z,\theta_0,h_0(w^T)) = R_t(w^t,\theta_0)[c - h_0(w^T)] + v_t,$$

and so

$$E[\rho_t(z,\theta_0,h_0(w^T))\,|\,w^t] = 0.$$

As in the general case, there is a simplifying transformation. Note that

$$H_t \equiv E[\partial\rho_t(z,\theta_0,h_0(w^T))/\partial h\,|\,w^T] = -R_t(w^t,\theta_0).$$

It will facilitate developing operational estimators if we do not evaluate R_t at θ_0, as in the transformation in Section 2. Instead we shall use the following transformation:

$$\begin{aligned}\lambda_t(z,\theta) &= \rho_t(z,\theta,h) - R_t(w^t,\theta)R_T^{-1}(w^T,\theta)\rho_T(z,\theta,h)\\ &= d_t(z,\theta) - R_t(w^t,\theta)R_T^{-1}(w^T,\theta)d_T(z,\theta).\end{aligned}$$

This transformation eliminates h, not just to first order as in Section 2. Versions of this transformation have been used by Wooldridge (1991) and Chamberlain (1992b).

Now we have sequential moment restrictions without having to deal with the nonparametric function h:

$$E[\lambda_t(z,\theta_0)\,|\,w^t]=0 \qquad (t=1,\ldots,T-1). \tag{3.2}$$

The variance bound for θ_0 is obtained by forward filtering:

$$\tilde{\lambda}_{T-1}(z,\theta)=\lambda_{T-1}(z,\theta),$$

$$\tilde{\lambda}_s(z,\theta)=\lambda_s(z,\theta)-\Gamma_{s,s+1}\tilde{\lambda}_{s+1}(z,\theta)$$

$$-\ldots-\Gamma_{s,T-1}\tilde{\lambda}_{T-1}(z,\theta) \qquad (s=T-2,\ldots,1),$$

where

$$\Gamma_{st}=E[\lambda_s(z,\theta_0)\tilde{\lambda}_t(z,\theta_0)\,|\,w^t]\big[E[\tilde{\lambda}_t^2(z,\theta_0)\,|\,w^t]\big]^{-1} \qquad (s<t).$$

Then we have

$$V_\theta=\left[\sum_{t=1}^{T-1}E\left[E\Big[\frac{\partial\tilde{\lambda}_t(z,\theta_0)}{\partial\theta}\,|\,w^t\Big]\big[E[\tilde{\lambda}_t^2(z,\theta_0)\,|\,w^t]\big]^{-1}E\Big[\frac{\partial\tilde{\lambda}_t(z,\theta_0)}{\partial\theta'}\,|\,w^t\Big]\right]\right]^{-1}.$$

This coincides with the variance bound in (2.5) because

$$\tilde{\lambda}_t(z,\theta_0)=\tilde{g}_t(z,\theta_0,h_0(w^T))$$

and

$$E[\partial\tilde{\lambda}_t(z,\theta_0)/\partial\theta'\,|\,w^t]=E[\partial\tilde{g}_t(z,\theta_0,h_0(w^T))/\partial\theta'\,|\,w^t].$$

The conditional moment restrictions in (3.2) lead directly to consistent, instrumental-variable estimators. We can choose a function M_t that maps the support of w^t into $\mathcal{R}^p$ $(t=1,\ldots,T-1)$. Then we can form the moment function

$$\psi(z,\theta)=\sum_{t=1}^{T-1}M_t(w^t)\lambda_t(z,\theta),$$

where ψ is $p\times 1$ (as is θ) and $E\psi(z,\theta_0)=0$. Hence, under suitable regularity conditions (see Hansen (1982)), the solution $\hat{\theta}$ to $\sum_{i=1}^n\psi(z_i,\theta)=0$ satisfies $\sqrt{n}(\hat{\theta}-\theta_0)\xrightarrow{D}N(0,\Lambda_\psi)$, where

$$\Lambda_\psi=\left[E\frac{\partial\psi(z,\theta_0)}{\partial\theta'}\right]^{-1}[E\psi(z,\theta_0)\psi'(z,\theta_0)]\left[E\frac{\partial\psi'(z,\theta_0)}{\partial\theta}\right]^{-1}.$$

An optimal instrumental-variable estimator has $\Lambda_\psi = V_\theta$. The solution in terms of the $\tilde{\lambda}_t$ is

$$\tilde{M}_t(w^t) = E\left[\frac{\partial\tilde{\lambda}_t(z,\theta_0)}{\partial\theta} \mid w^t\right]\left[E[\tilde{\lambda}_t^2(z,\theta_0) \mid w^t]\right]^{-1}.$$

Then we can recursively obtain $M_1 = \tilde{M}_1$ and

$$M_t = \tilde{M}_t - M_{t-1}\Gamma_{t-1,t} - \cdots - M_1\Gamma_{1t}$$

for $t = 2, \ldots, T-1$. These optimal instrumental variables involve various conditional expectation functions. They would have to be estimated using nonparametric techniques such as those developed by Robinson (1987, 1991) and Newey (1990, 1992).

An alternative approach uses a sequence of generalized method-of-moments (GMM) estimators based on an expanding set of instrumental variables, such as polynomials in w^t. Chamberlain (1992b) provides conditions on the sequence of instrumental variables such that the asymptotic variance of the GMM estimator based on k instrumental variables converges to V_θ as $k \to \infty$. Hahn (1991) shows how k should be chosen as a function of n in order for the sequence of GMM estimators based on $k(n)$ instrumental variables to have a limiting normal distribution with V_θ as the covariance matrix. This extends a result Newey (1992) obtained for the case in which the conditional moment restrictions all involve the same set of conditioning variables (i.e., $w^t = w$ for all t).

As for ϕ_0, evaluating $h_0 - H_T^{-1}\rho_T$ in (2.7) gives

$$h_0(w^T) - H_T^{-1}\rho_T(z,\theta_0,h_0(w^T)) = R_T^{-1}(w^T,\theta_0)d_T(z,\theta_0).$$

If θ_0 were known, we could use the following estimator of ϕ_0:

$$\hat{\phi} = \frac{1}{n}\sum_{i=1}^{n}\left[R_T^{-1}(w_i^T,\theta_0)d_T(z_i,\theta_0) - \sum_{t=1}^{T-1}A_t(w_i^t)\lambda_t(z_i,\theta_0)\right] \tag{3.3}$$
$$\to \phi_0 \text{ a.s.} \quad \text{as} \quad n \to \infty$$

since

$$E[R_T^{-1}(w_i^T,\theta_0)d_T(z_i,\theta_0)] = E(c_i) = \phi_0$$

and

$$E[\lambda_t(z_i,\theta_0)\,|\,w_i^t] = 0.$$

Then the asymptotic variance of $\hat{\phi}$ could be reduced to

$$\min_{A_1,\ldots,A_{T-1}} E\left[R_T^{-1}(w_i^T,\theta_0)d_T(z_i,\theta_0) - \phi_0 - \sum_{t=1}^{T-1} A_t(w_i^t)\lambda_t(z_i,\theta_0)\right]^2,$$

which gives the first term in (2.7).

Note that our analysis has been based on the restriction that $E(v_{it}\,|\,w_i^t) = 0$ in (3.1). It would be of interest to derive an efficiency bound under the stronger restriction that $E(v_{it}\,|\,w_i^t, c_i) = 0$. Ahn and Schmidt (1993) show that such an assumption can lead to additional moment restrictions on the observable variables.

4. IDENTIFICATION PROBLEMS

Suppose now that the random effect is not scalar, but has two (or more) components. We shall try to extend the transformation that was used in Section 3 to eliminate the random effect. It turns out, however, that it is not possible, in general, to construct such a transformation. This suggests there are identification problems in feedback models with a vector of random effects. We shall provide some examples to illustrate these problems.

Let the vector of random effects be denoted by $c_i' = (c_{1i}, \ldots, c_{Ji})$ and write the model as follows:

$$d(z_i, \theta_0) = R(w_i^T, \theta_0)c_i + v_i$$

$$E(v_{it} \,|\, w_i^t) = 0 \qquad (t = 1, \ldots, T),$$

where $d(z_i, \theta)$ is a $T \times 1$ vector, and $R(w_i^T, \theta)$ is a $T \times J$ matrix with t^{th} row equal to $(R_{t1}(w_i^t, \theta), \ldots, R_{tJ}(w_i^t, \theta))$.

We would like to construct a $(T - J) \times T$ transformation matrix of rank $T - J$ such that (dropping the i subscripts): (i) $Q(w^T, \theta)$ is upper-triangular (i.e., $Q_{ts}(w^T, \theta)$, the (t, s) element of $Q(w^T, \theta)$, is 0 for $t > s$); (ii) $Q_{ts}(w^T, \theta)$ depends only on w^s; (iii) $Q(w^T, \theta)R(w^T, \theta) = 0$. Then

$$Q(w^T, \theta_0)d(z, \theta_0) = Q(w^T, \theta_0)v \equiv \tilde{v}$$

with

$$\tilde{v}_t = \sum_{s=t}^{T} Q_{ts}(w^s, \theta_0)v_s$$

and

$$E(\tilde{v}_t \,|\, w^t) = 0 \qquad (t = 1, \ldots, T - J).$$

Define

$$\lambda_t(z, \theta) = \sum_{s=t}^{T} Q_{ts}(w^s, \theta)d_s(z, \theta);$$

then we have

$$E[\lambda_t(z, \theta_0) \,|\, w^t] = 0 \qquad (t = 1, \ldots, T - J).$$

So the problem reduces to a sequential set of conditional moment restrictions in which the nonparametric component has been eliminated.

There is no problem in constructing such a transformation when $J = \dim(c) = 1$ and $R_T(w^T, \theta) \equiv R_{T1}(w^T, \theta) \neq 0$ with probability one. For example, we can use

$$Q(w^T, \theta) = \begin{pmatrix} 1 & 0 & \ldots & 0 & -R_1(w^1, \theta)R_T^{-1}(w^T, \theta) \\ 0 & 1 & \ldots & 0 & -R_2(w^2, \theta)R_T^{-1}(w^T, \theta) \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -R_{T-1}(w^{T-1}, \theta)R_T^{-1}(w^T, \theta) \end{pmatrix},$$

which gives the transformation used in Section 3.

Satisfying the three requirements for Q is not, however, possible in general when $\dim(c) = 2$. Setting the last row of $Q(w^T, \theta)R(w^T, \theta)$ equal to zero gives

$$\sum_{s=T-2}^{T} Q_{T-2,s}(w^s, \theta)R_{s1}(w^s, \theta) = 0, \qquad \sum_{s=T-2}^{T} Q_{T-2,s}(w^s, \theta)R_{s2}(w^s, \theta) = 0,$$

and so both $Q_{T-2,T}(w^T, \theta)R_{T1}(w^T, \theta)$ and $Q_{T-2,T}(w^T, \theta)R_{T2}(w^T, \theta)$ must be functions of w^{T-1}, which implies that $R_{T1}(w^T, \theta)/R_{T2}(w^T, \theta)$ must be a function of w^{T-1}; but this would not be true in general.

The following example illustrates some of the identification issues:

$$y_{i1} = c_{1i} + c_{2i}x_{i1} + v_{i1}, \qquad E(v_{i1} \,|\, x_{i1}) = 0 \tag{4.1}$$

$$y_{i2} = \theta_0 + c_{1i} + c_{2i}x_{i2} + v_{i2}, \qquad E(v_{i2} \,|\, x_{i1}, x_{i2}) = 0. \tag{4.2}$$

Consider the following moment restrictions (dropping the i subscripts):

$$E(y_1 - y_2 + \theta_0 - c_2(x_1 - x_2) \,|\, x_1) = 0 \tag{4.3}$$

$$E(y_2 - \theta_0 - c_1 - c_2x_2 \,|\, x_1, x_2) = 0. \tag{4.4}$$

These are equivalent to the moment restrictions in (4.1) and (4.2) because $E(v_2 \,|\, x_1, x_2) = 0$ implies that $E(v_2 \,|\, x_1) = 0$, which, combined with $E(v_1 - v_2 \,|\, x_1) = 0$, implies that $E(v_1 \,|\, x_1) = 0$. Since $E(c_1 \,|\, x_1, x_2)$ is unrestricted, only (4.3) is relevant for the identification of θ_0 and a conditional mean of c_2.

Suppose that x_1 and x_2 are binary variables, equal to 0 or 1. Then identification must be based on the following two equations:

$$E(y_1 - y_2 \,|\, x_1 = 0) = -\theta_0 - P(x_2 = 1 \,|\, x_1 = 0)E(c_2 \,|\, x_1 = 0, x_2 = 1)$$

$$E(y_1 - y_2 \,|\, x_1 = 1) = -\theta_0 + P(x_2 = 0 \,|\, x_1 = 1)E(c_2 \,|\, x_1 = 1, x_2 = 0).$$

These two equations are not sufficient to identify any of the three parameters θ_0, $E(c_2 \,|\, x_1 = 0, x_2 = 1)$, or $E(c_2 \,|\, x_1 = 1, x_2 = 0)$. We can, however, identify a certain convex combination of the c_2 conditional means:

$$[p_{1|0}E(c_2 \,|\, x_1 = 0, x_2 = 1) + p_{0|1}E(c_2|x_1 = 1, x_2 = 0)]/(p_{1|0} + p_{0|1})$$

$$= [E(y_1 - y_2 \,|\, x_1 = 1) - E(y_1 - y_2 \,|\, x_1 = 0)]/(p_{1|0} + p_{0|1}),$$

where $p_{a|b} \equiv P(x_2 = a \,|\, x_1 = b)$. The corresponding estimator is the instrumental-variable estimator that uses $(1, x_1)$ as instrumental variables with $y_1 - y_2$ as the dependent variable and with $(1, x_1 - x_2)$ as the explanatory variables.

Next we shall extend this example to show that the failure of identification for θ_0 is quite general. We shall also gain some insight into when it is possible to identify some convex combination of the c_2 conditional means.

Consider the following model:

$$y_{it} = \theta_0' r_{it} + c_{1i} + c_{2i} x_{it} + v_{it} \tag{4.5}$$

$$E(v_{it} \,|\, w_i^t) = 0 \qquad (t = 1, \ldots, T).$$

Here r_{it} is $p \times 1$, x_{it} is scalar, $w_{it}' = (r_{it}', x_{it})$, and $w_i^t = (w_{i1}, \ldots, w_{it})$; r_{it} can include lagged values of the outcome variable as well as period effects.

The moment restrictions in (4.5) have the following implications (dropping the i subscripts):

$$E[y_t - y_{t+1} - \theta_0'(r_t - r_{t+1}) - c_2(x_t - x_{t+1}) \,|\, w^t] = 0 \qquad (t = 1, \ldots, T-1) \tag{4.6}$$

$$E[y_T - \theta_0' r_T - c_1 - c_2 x_T \,|\, w^T] = 0. \tag{4.7}$$

Note that (4.6) and (4.7) imply that

$$E(v_T \,|\, w^{T-1}) = 0, \qquad E(v_{T-1} - v_T \,|\, w^{T-1}) = 0,$$

and so $E(v_{T-1} \,|\, w^{T-1}) = 0$. Continuing recursively in this fashion shows that the moment restrictions in (4.6) and (4.7) are equivalent to the moment restrictions in (4.5). In addition, since $E(c_1 \,|\, w^T)$ is unrestricted, only (4.6) is relevant for the identification of θ_0.

PROPOSITION 1. Suppose that x_t has finite support $(t = 1, \ldots, T)$ and that for any point a in the support of w^t, the distribution of x_{t+1} conditional on $w^t = a$ is not degenerate $(t \leq T-1)$. Then θ_0 is not identified.

PROOF. Let the support of x_t be $\{\delta_{1t}, \ldots, \delta_{L_t t}\}$. We shall show that given an arbitrary point $\tilde{\theta} \in \mathcal{R}^p$, we can construct $E(\tilde{c}_2 \,|\, w^T)$ such that (4.6) is satisfied with $\tilde{\theta}$ replacing θ_0 and with $E(\tilde{c}_2 \,|\, w^T)$ replacing $E(c_2 \,|\, w^T)$.

Set $E(\tilde{c}_2 \,|\, w^1) = 0$. We shall recursively construct $E(\tilde{c}_2 \,|\, w^{t+1})$ given $E(\tilde{c}_2 \,|\, w^t)$. We can write (4.6) as

$$\begin{aligned} &E[y_t - y_{t+1} - \tilde{\theta}'(r_t - r_{t+1}) \,|\, w^t] \\ &\quad = \sum_{j=1}^{L_{t+1}} P(x_{t+1} = \delta_{j,t+1} \,|\, w^t)(x_t - \delta_{j,t+1}) E(\tilde{c}_2 \,|\, w^t, x_{t+1} = \delta_{j,t+1}). \end{aligned} \tag{4.8}$$

There are at least two distinct values for x_{t+1} that have positive probability conditional on w^t; denote these values by $\delta_{k,t+1}$ and $\delta_{l,t+1}$ (where k and l depend upon w^t). Let

$$E(\tilde{c}_2 \,|\, w^t, x_{t+1} = \delta_{j,t+1}) = 0 \quad \text{for} \quad j \neq k \text{ or } l$$

and

$$\begin{pmatrix} E(\tilde{c}_2 \,|\, w^t, x_{t+1} = \delta_{k,t+1}) \\ E(\tilde{c}_2 \,|\, w^t, x_{t+1} = \delta_{l,t+1}) \end{pmatrix}$$

$$= \begin{pmatrix} P(x_{t+1} = \delta_{k,t+1} \,|\, w^t)(x_t - \delta_{k,t+1}) & P(x_{t+1} = \delta_{l,t+1} \,|\, w^t)(x_t - \delta_{l,t+1}) \\ P(x_{t+1} = \delta_{k,t+1} \,|\, w^t) & P(x_{t+1} = \delta_{l,t+1} \,|\, w^t) \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} E[y_t - y_{t+1} - \tilde{\theta}'(r_t - r_{t+1}) \,|\, w^t] \\ E(\tilde{c}_2 \,|\, w^t) \end{pmatrix}.$$

The matrix is nonsingular because $P(x_{t+1} = \delta_{k,t+1} \,|\, w^t)$ and $P(x_{t+1} = \delta_{l,t+1} \,|\, w^t)$ are positive and $\delta_{k,t+1} \neq \delta_{l,t+1}$.

Now we have constructed $E(\tilde{c}_2 \,|\, w^t, x_{t+1})$ such that (4.8) is satisfied and

$$E(\tilde{c}_2 \,|\, w^t) = E[E(\tilde{c}_2 \,|\, w^t, x_{t+1}) \,|\, w^t].$$

Complete the construction of $E(\tilde{c}_2 \,|\, w^{t+1})$ by setting

$$E(\tilde{c}_2 \,|\, w^t, r_{t+1}, x_{t+1}) = E(\tilde{c}_2 \,|\, w^t, x_{t+1}).$$

Then

$$E(\tilde{c}_2 \,|\, w^t) = E[E(\tilde{c}_2 \,|\, w^{t+1}) \,|\, w^t].$$

Continue in this way until finally $E(\tilde{c}_2 \,|\, w^T)$ is constructed from $E(\tilde{c}_2 \,|\, w^{T-1})$. Then $\tilde{\theta}$ and $E(\tilde{c}_2 \,|\, w^T)$ satisfy (4.8) for $t = 1, \ldots, T-1$. Q.E.D.

Even if θ_0 were identified, we may not be able to identify a convex combination of conditional means of c_2. The problem is that $(x_t - \delta_{j,t+1})$ in (4.8) generally changes sign as we run through the support of x_{t+1}. Conditioning on $w^t = a$ provides a convex combination of $E(c_2 \,|\, w^t = a, x_{t+1} = \delta_{j,t+1})$ only if

$$P(x_t \leq x_{t+1} \,|\, w^t = a) = 1 \quad \text{or} \quad P(x_t \geq x_{t+1} \,|\, w^t = a) = 1. \tag{4.9}$$

Our first example had $x_t = 0$ or 1 for $t = 1, 2$. Hence (4.9) was satisfied. But if, for example, the support of x_2 conditional on $x_1 = 0$ is $\{-1, 1\}$, and the support of x_2 conditional on $x_1 = 1$ is $\{0, 2\}$, then we would not be able to identify a convex combination of the $E(c_2 \,|\, x_1, x_2)$ values. More generally, the identification prospects diminish as the support of x_t becomes richer.

APPENDIX

A.1 Proof of Theorem 1

The conditional moment restrictions in Condition (C) are, in the multinomial case, equivalent to a finite set of unconditional moment restrictions. We show this by setting up dummy indicator variables that pick out the support points of the distribution of w^t. Let $X_t = \{\tau_{t1}, \ldots, \tau_{tl_t}\}$ be the support of the distribution of w^t. Define $d_{tj}\colon X_t \to \mathcal{R}$ by $d_{tj}(a) = 1$ if $a = \tau_{tj}$ and $= 0$ otherwise. Define

$$\psi(z,\theta,\lambda,\phi) = \begin{pmatrix} \psi_1(z,\theta,\lambda) \\ \vdots \\ \psi_T(z,\theta,\lambda) \\ \psi_{T+1}(z,\lambda,\phi) \end{pmatrix} = \begin{pmatrix} d_1(w^1)\rho_1(z,\theta,\sum_{j=1}^{l_T}\lambda_j d_{Tj}(w^T)) \\ \vdots \\ d_T(w^T)\rho_T(z,\theta,\sum_{j=1}^{l_T}\lambda_j d_{Tj}(w^T)) \\ \phi - \sum_{j=1}^{l_T}\lambda_j d_{Tj}(w^T) \end{pmatrix},$$

where $d_t' = (d_{t1}, \ldots, d_{tl_t})$. Then for $t = 1, \ldots, T$,[4]

$$E[\psi_t(z,\theta,\lambda)] = E\big[d_t(w^t)E[\rho_t(z,\theta,\sum_{j=1}^{l_T}\lambda_j d_{Tj}(w^T))\,|\,w^t]\big]$$

$$= \begin{pmatrix} P\{w^t = \tau_{t1}\}E[\rho_t(z,\theta,\sum_{j=1}^{l_T}\lambda_j d_{Tj}(w^T))\,|\,w^t = \tau_{t1}] \\ \vdots \\ P\{w^t = \tau_{tl_t}\}E[\rho_t(z,\theta,\sum_{j=1}^{l_T}\lambda_j d_{Tj}(w^T))\,|\,w^t = \tau_{tl_t}] \end{pmatrix},$$

and

$$E[\psi_{T+1}(z,\lambda,\phi)] = \phi - \sum_{j=1}^{l_T} P\{w^T = \tau_{Tj}\}\lambda_j.$$

So if we set $h(\tau_{Tj}) = \lambda_j$, then $E[\psi_t(z,\theta,\lambda)] = 0$ implies that

$$E[\rho_t(z,\theta,h(w^T))\,|\,w^t] = 0,$$

and $E[\psi_{T+1}(z,\lambda,\phi)] = 0$ implies that

$$\phi = E[h(w^T)].$$

In addition, with $\lambda_{0j} = h_0(\tau_{Tj})$, Condition (C) implies that

$$E[\psi(z, \theta_0, \lambda_0, \phi_0)] = 0.$$

It is shown in Chamberlain (1987, Lemma 2) that, in the multinomial case, the Fisher information bound for μ_0 under the restriction that $E[\psi(z, \mu_0)] = 0$ is

$$J_\psi \equiv E\left[\frac{\partial \psi(z, \mu_0)}{\partial \mu'}\right]' [E\psi(z, \mu_0)\psi'(z, \mu_0)]^{-1} E\left[\frac{\partial \psi(z, \mu_0)}{\partial \mu'}\right].$$

This lemma requires that $E[\psi(z, \mu_0)\psi'(z, \mu_0)]$ is nonsingular and that $E[\partial \psi(z, \mu_0)/\partial \mu']$ has full column rank; (PD) implies that these conditions hold.

Let $\mu_0' = (\theta_0', \lambda_0', \phi_0)$. We shall construct an equivalent moment function, $\tilde{\psi}(z, \mu) = A\psi(z, \mu)$, where the matrix A is nonsingular, so that $J_{\tilde{\psi}} = J_\psi$.

Define A_1 such that

$$A_1 \psi = \begin{pmatrix} \psi_1 - B_1 \psi_T \\ \psi_2 \\ \vdots \\ \psi_T \\ \psi_{T+1} \end{pmatrix},$$

where ψ is evaluated at (z, μ) with $\mu' = (\theta', \lambda', \phi)$. Define B_1^* to be a matrix of 0's and 1's such that $B_1^* d_T(w^T) = d_1(w^1)$ and let

$$B_1 = B_1^* \text{diag}\{H_1(\tau_{T1}) H_T^{-1}(\tau_{T1}), \ldots, H_1(\tau_{Tl_T}) H_T^{-1}(\tau_{Tl_T})\},$$

where $H_t(a) \equiv E[\partial \rho_t(z, \theta_0, h_0(w^T))/\partial h \,|\, w^T = a]$. Then

$$\begin{aligned} B_1 d_T(w^T)\rho_T &= B_1^* \begin{pmatrix} H_1(\tau_{T1}) H_T^{-1}(\tau_{T1}) d_{T1}(w^T) \\ \vdots \\ H_1(\tau_{Tl_T}) H_T^{-1}(\tau_{Tl_T}) d_{Tl_T}(w^T) \end{pmatrix} \rho_T \\ &= B_1^* d_T(w^T) H_1(w^T) H_T^{-1}(w^T) \rho_T \\ &= d_1(w^1) H_1(w^T) H_T^{-1}(w^T) \rho_T, \end{aligned}$$

and so

$$\psi_1 - B_1\psi_T = d_1(w^1)[\rho_1 - H_1(w^T)H_T^{-1}(w^T)\rho_T]$$
$$= d_1(w^1)g_1,$$

where ρ_1, ρ_T, and g_1 are evaluated at $(z, \theta, \sum_{j=1}^{l_T} \lambda_j d_{Tj}(w^T))$.

With $A_2, \ldots, A_{T-1}$ defined in a similar fashion, we have

$$A_{T-1} \times \cdots \times A_1\psi = \begin{pmatrix} d_1 g_1 \\ \vdots \\ d_{T-1}g_{T-1} \\ d_T\rho_T \\ \psi_{T+1} \end{pmatrix} \equiv \psi^*.$$

Note that $A_{T-1} \times \cdots \times A_1$ is nonsingular, since it is upper triangular with ones on the diagonal.

It is shown in Chamberlain (1992b, p. 24) that there is a nonsingular matrix A^* such that

$$\tilde{\psi} \equiv A^*\psi^* = \begin{pmatrix} d_1\tilde{g}_1 \\ \vdots \\ d_{T-1}\tilde{g}_{T-1} \\ d_T\rho_T \\ \psi_{T+1} \end{pmatrix},$$

where $\tilde{g}_t$ is defined in (2.2).

Since $E[\partial\tilde{g}_t(z, \theta_0, h_0(w^T))/\partial h \,|\, w^t] = 0$, we have

$$E[\partial\tilde{\psi}(z, \mu_0)/\partial\mu'] = \begin{pmatrix} \Gamma_1 & 0 \\ \Gamma_2 & \Gamma_3 \end{pmatrix}$$

where

$$\tilde{G}_t(a) = E[\partial\tilde{g}_t(z, \theta_0, h_0(w^T))/\partial\theta' \,|\, w^t = a],$$

$$N_t = E[d_t(w^t)\tilde{G}_t(w^t)] = \begin{pmatrix} P(\tau_{t1})\tilde{G}_t(\tau_{t1}) \\ \vdots \\ P(\tau_{tl_t})\tilde{G}_t(\tau_{tl_t}) \end{pmatrix},$$

$$\Gamma_1 = \begin{pmatrix} N_1 \\ \vdots \\ N_{T-1} \end{pmatrix},$$

with $P(\tau_{tj}) = P\{w^t = \tau_{tj}\}$;

$$D_T(a) = E[\partial \rho_T(z, \theta_0, h_0(w^T))/\partial \theta' \,|\, w^T = a],$$

$$M = E[d_T(w^T) D_T(w^T)] = \begin{pmatrix} P(\tau_{T1}) D_T(\tau_{T1}) \\ \vdots \\ P(\tau_{Tl_T}) D_T(\tau_{Tl_T}) \end{pmatrix},$$

$$\Gamma_2 = \begin{pmatrix} M \\ 0 \end{pmatrix};$$

$$c' = -E[d_T'(w^T)] = -[P(\tau_{T1}), \ldots, P(\tau_{Tl_T})],$$

$$Q = E\big[E[\partial \rho_T(z, \theta_0, h_0(w^T))/\partial h \,|\, w^T] d_T(w^T) d_T'(w^T)]$$

$$= \text{diag}\{P(\tau_{T1}) H_T(\tau_{T1}), \ldots, P(\tau_{Tl_T}) H_T(\tau_{Tl_T})\},$$

$$\Gamma_3 = \begin{pmatrix} Q & 0 \\ c' & 1 \end{pmatrix}.$$

Next we shall derive $E(\tilde{\psi}\tilde{\psi}')$. $\tilde{g}_t$ and ρ_T are evaluated at $(z, \theta_0, h_0(w^T))$, and h_0 is evaluated at w^T. We shall also simplify notation by letting $E(\,\cdot\,|\,\tau_{tj})$ denote $E(\,\cdot\,|\,w^t = \tau_{tj})$; for example, $E(\tilde{g}_t h_0 \,|\, \tau_{tj})$ denotes $E(\tilde{g}_t h_0 \,|\, w^t = \tau_{tj})$.

$$E[\tilde{\psi}(z, \mu_0) \tilde{\psi}'(z, \mu_0)] = \Psi = \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix}$$

where

$$\Psi_{11} = \text{diag}\{P(\tau_{11}) \tilde{\Sigma}_1(\tau_{11}), \ldots, P(\tau_{1l_1}) \tilde{\Sigma}_1(\tau_{1l_1}), \ldots,$$
$$P(\tau_{T-1,1}) \tilde{\Sigma}_{T-1}(\tau_{T-1,1}), \ldots, P(\tau_{T-1,l_{T-1}}) \tilde{\Sigma}_{T-1}(\tau_{T-1,l_{T-1}})\}$$

(since $E(\tilde{g}_s \tilde{g}_t \,|\, w^t) = 0$ for $s < t$), where $\tilde{\Sigma}_t(a) = E(\tilde{g}_t^2 \,|\, w^t = a)$. Order the support points so that $\tau_{t+1,1}, \ldots, \tau_{t+1,k_1}$ each has τ_{t1} as the first set of components, $\tau_{t+1,k_1+1}, \ldots, \tau_{t+1,k_2}$ each has τ_{t2} as the first set of components, and so on; then

$$\Psi_{12} = \Psi_{21}' = \begin{pmatrix} C_1 & F_1 \\ \vdots & \vdots \\ C_{T-1} & F_{T-1} \end{pmatrix}$$

where

$$C_t = \begin{pmatrix} q'_{t1} & 0 & \dots & 0 \\ 0 & q'_{t2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q'_{tl_t} \end{pmatrix}, \qquad F_t = -\begin{pmatrix} P(\tau_{t1})E(\tilde{g}_t h_0 \mid \tau_{t1}) \\ \vdots \\ P(\tau_{tl_t})E(\tilde{g}_t h_0 \mid \tau_{tl_t}) \end{pmatrix},$$

with q_{tj} a column vector with elements of the form $P(\tau_{Tl})E(\tilde{g}_t\rho_T \mid \tau_{Tl})$ for l such that τ_{Tl} has τ_{tj} for its first set of components—which we denote by $\tau_{tj} \in \tau_{Tl}$, so that $q_{tj} = (P(\tau_{Tl})E(\tilde{g}_t\rho_T \mid \tau_{Tl}))_{\{\,l:\tau_{tj}\in\tau_{Tl}\}}$.

$$\Psi_{22} = \begin{pmatrix} \text{diag}\{P(\tau_{Tj})E(\rho_T^2 \mid \tau_{Tj})\}_{j=1}^{l_T} & 0 \\ 0 & \text{Var}(h_0) \end{pmatrix}.$$

Let

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} \Gamma_1 & 0 \\ \Gamma_2 & \Gamma_3 \end{pmatrix}' \Psi^{-1} \begin{pmatrix} \Gamma_1 & 0 \\ \Gamma_2 & \Gamma_3 \end{pmatrix}, \tag{A.1}$$

and let Ω^{jk} denote the (j,k) block of Ω^{-1}. Then (see Chamberlain (1992b, lemma 1))

$$\begin{aligned} V_\theta^{-1} = (\Omega^{11})^{-1} &= \Gamma_1'\Psi_{11}^{-1}\Gamma_1 \\ &= \sum_{t=1}^{T-1} N_t'\text{diag}\{P^{-1}(\tau_{tj})\tilde{\Sigma}_t^{-1}(\tau_{tj})\}_{j=1}^{l_t} N_t \\ &= \sum_{t=1}^{T-1}\sum_{j=1}^{l_t} P(\tau_{tj})\tilde{G}_t'(\tau_{tj})\tilde{\Sigma}_t^{-1}(\tau_{tj})\tilde{G}_t(\tau_{tj}) \\ &= \sum_{t=1}^{T-1} E(\tilde{G}_t'\tilde{\Sigma}_t^{-1}\tilde{G}_t), \end{aligned}$$

where $\tilde{G}_t = E(\partial\tilde{g}_t/\partial\theta' \mid w^t)$ and $\tilde{\Sigma}_t = E(\tilde{g}_t^2 \mid w^t)$ with $\tilde{g}_t$ evaluated at $(z, \theta_0, h_0(w^T))$.

Let Ψ^{jk} denote the (j,k) block of Ψ^{-1} $(j,k = 1,2)$. We shall derive Ψ^{22}.

$$\Psi^{22} = (\Psi_{22} - \Psi_{21}\Psi_{11}^{-1}\Psi_{12})^{-1};$$

$$\Psi_{21}\Psi_{11}^{-1}\Psi_{12} = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}$$

where

$$W_{11} = \sum_{t=1}^{T-1} \operatorname{diag}\{P^{-1}(\tau_{t1})\tilde{\Sigma}_t^{-1}(\tau_{t1})q_{t1}q'_{t1}, \ldots, P^{-1}(\tau_{tl_t})\tilde{\Sigma}_t^{-1}(\tau_{tl_t})q_{tl_t}q'_{tl_t}\},$$

$$W_{12} = W'_{21} = -\sum_{t=1}^{T-1} \begin{pmatrix} q_{t1}\tilde{\Sigma}_t^{-1}(\tau_{t1})E(\tilde{g}_t h_0 \mid \tau_{t1}) \\ \vdots \\ q_{tl_t}\tilde{\Sigma}_t^{-1}(\tau_{tl_t})E(\tilde{g}_t h_0 \mid \tau_{tl_t}) \end{pmatrix},$$

$$W_{22} = \sum_{t=1}^{T-1} E\big[[E(\tilde{g}_t h_0 \mid w^t)]^2 \tilde{\Sigma}_t^{-1}\big].$$

Let

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = \Psi_{22} - \Psi_{21}\Psi_{11}^{-1}\Psi_{12}.$$

Then

$$M_{11} = \operatorname{diag}\{P(\tau_{Tj})E(\rho_T^2 \mid \tau_{Tj})\}_{j=1}^{l_T} - W_{11},$$

$$M_{12} = M'_{21} = -W_{12},$$

$$M_{22} = \operatorname{Var}(h_0) - \sum_{t=1}^{T-1} E\big[[E(\tilde{g}_t h_0 \mid w^t)]^2 \tilde{\Sigma}_t^{-1}\big].$$

If a and b are functions mapping the support of w^T into $\mathcal{R}$, and if

$$\underline{a}' = (a(\tau_{T1}), \ldots, a(\tau_{Tl_T}))$$

$$\underline{b}' = (b(\tau_{T1}), \ldots, b(\tau_{Tl_T})),$$

then

$$\underline{a}' M_{11} \underline{b} = E[a(w^T)b(w^T)\rho_T^2] - \sum_{t=1}^{T-1} E\big[E[a(w^T)\tilde{g}_t\rho_T \mid w^t]\tilde{\Sigma}_t^{-1} E[b(w^T)\tilde{g}_t\rho_T \mid w^t]\big]$$

and

$$\underline{a}' M_{12} = \sum_{t=1}^{T-1} E\big[E(\tilde{g}_t h_0 \mid w^t) E[a(w^T)\tilde{g}_t\rho_T \mid w^t]\tilde{\Sigma}_t^{-1}\big].$$

Next we shall derive Ω_{22}^{-1}.

$$\begin{aligned}\Omega_{22} &= \Gamma_3' \Psi^{22} \Gamma_3 \\ &= \begin{pmatrix} Q & c \\ 0 & 1 \end{pmatrix} \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}^{-1} \begin{pmatrix} Q & 0 \\ c' & 1 \end{pmatrix},\end{aligned}$$

$$\Omega_{22}^{-1} = \begin{pmatrix} Q^{-1} & 0 \\ -c'Q^{-1} & 1 \end{pmatrix} \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} Q^{-1} & -Q^{-1}c \\ 0 & 1 \end{pmatrix},$$

$$(\Omega_{22}^{-1})_{22} = M_{22} + c'Q^{-1}M_{11}Q^{-1}c - M_{21}Q^{-1}c - c'Q^{-1}M_{12},$$

where $(\Omega_{22}^{-1})_{22}$ denotes the element of Ω_{22}^{-1} in the lower-right corner.

$$c'Q^{-1} = -[H_T^{-1}(\tau_{T1}), \ldots, H_T^{-1}(\tau_{Tl_T})],$$

$$\begin{aligned}c'Q^{-1}M_{11}Q^{-1}c = E(&H_T^{-2}\rho_T^2) \\ &- \sum_{t=1}^{T-1} E\big[[E(H_T^{-1}\tilde{g}_t\rho_T \,|\, w^t)]^2 \tilde{\Sigma}_t^{-1}\big],\end{aligned}$$

$$-c'Q^{-1}M_{12} = \sum_{t=1}^{T-1} E\big[E(\tilde{g}_t h_0 \,|\, w^t) E(H_T^{-1}\tilde{g}_t\rho_T \,|\, w^t)\tilde{\Sigma}_t^{-1}\big],$$

where $H_T = H_T(w^T)$. Hence

$$(\Omega_{22}^{-1})_{22} = \mathrm{Var}\Big[h_0 - H_T^{-1}\rho_T - \sum_{t=1}^{T-1} \Lambda_t \tilde{g}_t\Big]$$

where

$$\Lambda_t = E[(h_0 - H_T^{-1}\rho_T)\tilde{g}_t \,|\, w^t]\tilde{\Sigma}_t^{-1}.$$

Next we shall derive Ω^{22}.

$$\Omega^{22} = \Omega_{22}^{-1} + \Omega_{22}^{-1}\Omega_{21}\Omega^{11}\Omega_{12}\Omega_{22}^{-1}$$

where

$$\begin{aligned}\Omega_{22}^{-1}\Omega_{21} &= (\Gamma_3'\Psi^{22}\Gamma_3)^{-1}(\Gamma_3'\Psi^{21}\Gamma_1 + \Gamma_3'\Psi^{22}\Gamma_2) \\ &= -\Gamma_3^{-1}\Psi_{21}\Psi_{11}^{-1}\Gamma_1 + \Gamma_3^{-1}\Gamma_2;\end{aligned}$$

$$-\Gamma_3^{-1}\Psi_{21}\Psi_{11}^{-1}\Gamma_1 = -\sum_{t=1}^{T-1}\begin{pmatrix} Q^{-1} & 0 \\ -c'Q^{-1} & 1 \end{pmatrix}\begin{pmatrix} q_{t1}\tilde{\Sigma}_t^{-1}(\tau_{t1})\tilde{G}_t(\tau_{t1}) \\ \vdots \\ q_{tl_t}\tilde{\Sigma}_t^{-1}(\tau_{tl_t})\tilde{G}_t(\tau_{tl_t}) \\ -E\big[E(\tilde{g}_t h_0 \mid w^t)\tilde{\Sigma}_t^{-1}\tilde{G}_t\big] \end{pmatrix}$$

$$= \begin{pmatrix} L \\ \sum_{t=1}^{T-1} E\big[E[(h_0 - H_T^{-1}\rho_T)\tilde{g}_t \mid w^t]\tilde{\Sigma}_t^{-1}\tilde{G}_t\big] \end{pmatrix}$$

(we shall not need to evaluate L);

$$\Gamma_3^{-1}\Gamma_2 = \begin{pmatrix} Q^{-1} & 0 \\ -c'Q^{-1} & 1 \end{pmatrix}\begin{pmatrix} M \\ 0 \end{pmatrix} = \begin{pmatrix} Q^{-1}M \\ -c'Q^{-1}M \end{pmatrix};$$

$$-c'Q^{-1}M = E(H_T^{-1}D_T),$$

where $D_T = D_T(w^T)$.

Since $\Omega^{11} = V_\theta$, we have

$$V_\phi = (\Omega^{22})_{22} = \mathrm{Var}\Big[h_0 - H_T^{-1}\rho_T - \sum_{t=1}^{T-1}\Lambda_t\tilde{g}_t\Big] + KV_\theta K'$$

where

$$K = \sum_{t=1}^{T-1} E(\Lambda_t\tilde{G}_t) + E(H_T^{-1}D_T).$$

Finally, we shall obtain $V_{\phi\theta}$.

$$\Omega^{21} = -\Omega_{22}^{-1}\Omega_{21}\Omega^{11}.$$

Hence

$$V_{\phi\theta} = V_{\theta\phi}' = -KV_\theta. \qquad \text{Q.E.D.}$$

A.2 The Minimax Bound for the General Case

We have shown that V_0 is the variance bound in the multinomial case. Since we can construct such a multinomial distribution in any neighborhood of a general distribution P, we can show that V_0 is the local minimax bound in the general case.

We shall need to consider neighborhoods of (P, θ_0, h_0). The neighborhoods of P are defined as follows: let $\mathcal{D}$ be the set of all probability measures on the Borel subsets of $\mathcal{R}^m$; we define the basic neighborhoods of P to be sets of the form

$$\{ Q \in \mathcal{D}\colon |\int f_j \, dQ - \int f_j \, dP| < \epsilon_j, \qquad j = 1, \ldots, k \},$$

where $\epsilon_j > 0$, k is some integer, and the $f_j\colon \mathcal{R}^m \to \mathcal{R}$ are measurable functions such that $\int |f_j| \, dP < \infty$. An arbitrary neighborhood of P is formed by taking unions of sets of this form.

The neighborhoods of h_0 are assumed to satisfy the following property: given any neighborhood Ψ of h_0 and given a finite subset $\{a_1, \ldots, a_k\}$ of the P-support of w^T, there is an $\epsilon > 0$ such that if $\zeta_j \in \mathcal{R}$ and $|\zeta_j| < \epsilon$, then there is an $h \in \Psi$ such that $h(a_j) = h_0(a_j) + \zeta_j$ for $j = 1, \ldots, k$. (Sup-norm neighborhoods will do: if Ψ consists of the measurable h mapping the P-support of w^T into $\mathcal{R}$ such that $\|h - h_0\| < \bar{\epsilon}$, then if $\epsilon < \bar{\epsilon}$ we can set $h(a_j) = h_0(a_j) + \zeta_j$ and $h(a) = h_0(a)$ for $a \notin \{a_1, \ldots, a_k\}$.) Then, with Euclidean neighborhoods for θ_0, the neighborhoods of (P, θ_0, h_0) are formed using the product topology.

We shall use the following class L of loss functions: $l \in L$ if for all $\alpha, \beta \in \mathcal{R}$,

(i) $\quad l(\alpha) = l(|\alpha|)$;

(ii) $\quad |\alpha| \leq |\beta|$ implies $l(\alpha) \leq l(\beta)$;

(iii) $\quad \int_{-\infty}^{\infty} l(\alpha) \exp(-\frac{1}{2}\lambda\alpha^2) \, d\alpha < \infty$ for $\lambda > 0$;

(iv) $\quad l(0) = 0$.

THEOREM 2. Suppose that (P, θ_0, h_0) satisfies Conditions (C) and (PD). Let Ψ be any neighborhood of (P, θ_0, h_0) and let Γ be the subset of Ψ such that Conditions (C) and (PD) are satisfied for all (Q, θ, h) in Γ. Define $\phi = E_Q[h(w^T)]$. Let α_s be the s^{th} component of $\alpha' \equiv (\theta', \phi)$ and let $T_n(z_1, \ldots, z_n)$ be any (measurable) estimator of α_s. Then for any loss function $l \in L$, we have

$$\liminf_{n\to\infty} \sup_{(Q,\theta,h)\in\Gamma} E_Q\{l[\sqrt{n}(T_n - \alpha_s)]\}$$
$$\geq (2\pi)^{-1/2} \int_{-\infty}^{\infty} l(\sigma_0 u) \exp(-\tfrac{1}{2}u^2)\, du,$$

where σ_0^2 is the (s, s) element of V_0.

The proof of Theorem 2 is similar to that of Theorem 1 in Chamberlain (1992a). It relies on the following auxiliary results.

LEMMA 1. Let z be a m-dimensional random vector; let w_t be a set of components of z and define $w^t = (w_1, \ldots, w_t)$ for $t = 1, \ldots, T$. Then given a probability measure P defined on the Borel subsets of R^m and a measurable function $f\colon \mathcal{R}^m \to \mathcal{R}^u$ with $E_P\|f\| < \infty$, there exists a probability measure M, with support a finite subset of the support of P, such that

$$E_M(f) = E_P(f), \tag{A.2}$$

and for each point a in the M-support of w^t

$$E_M(f \mid w^t = a) = E_P(f \mid w^t = a); \tag{A.3}$$

furthermore, given any Borel set A with $P\{w^t \in A\} = 1$, the M-support of w^t can be chosen to be a subset of A $(t = 1, \ldots, T)$.

PROOF. For a in the P-support of w^t, define $q_t(a) = E_P(f \mid w^t = a)$. From Chamberlain (1992a, Lemma A1), given a Borel set A with $P\{w^1 \in A\} = 1$, we can construct a probability measure M_1 with support a finite subset of A such that $E_{M_1}(q_1) = E_P[q_1(w^1)]$. Given M_t and a point a in the support of M_t, construct the probability measure $M_{t+1}(dw_{t+1} \mid w^t = a)$ with finite support such that

$$E_{M_{t+1}}(q_{t+1} \mid w^t = a) = E_P[q_{t+1}(w^{t+1}) \mid w^t = a].$$

Then define the probability measure M_{t+1} by $M_{t+1}(dw^{t+1}) = M_{t+1}(dw_{t+1} \mid w^t)M_t(dw^t)$. Given a Borel set A with $P\{w^{t+1} \in A\} = 1$, we can choose the support of

$$M_{t+1}(dw_{t+1} \mid w^t = a)$$

so that the support of M_{t+1} is a subset of A. (See the proof of Lemma 1 in Chamberlain (1992a).) In this way we construct $M_2, \ldots, M_T$.

Given a point a in the support of M_T, construct the probability measure $M(dz \mid w^T = a)$ with finite support such that

$$E_M(f \mid w^T = a) = E_P(f \mid w^T = a). \tag{A.4}$$

Then define the probability measure M by $M(dz) = M(dz \mid w^T)M_T(dw^T)$. We can choose the support of $M(dz \mid w^T = a)$ such that the support of M is a subset of the support of P.

Now check that M satisfies (A.2) and (A.3). It follows from (A.4) that (A.3) holds for $t = T$. Suppose that (A.3) holds for $s = T, \ldots, t+1$; then for each point a in the M-support of w^t,

$$\begin{aligned} E_M(f \mid w^t = a) &= E_M[E_M(f \mid w^{t+1}) \mid w^t = a] \\ &= E_M[E_P(f \mid w^{t+1}) \mid w^t = a] \\ &= E_{M_{t+1}}(q_{t+1} \mid w^t = a) \\ &= E_P[q_{t+1}(w^{t+1}) \mid w^t = a] \\ &= E_P[E_P(f \mid w^{t+1}) \mid w^t = a] \\ &= E_P(f \mid w^t = a). \end{aligned}$$

Hence (A.3) holds for $t = 1, \ldots, T$. As for (A.2),

$$\begin{aligned} E_M(f) &= E_M[E_M(f \mid w^1)] = E_M[E_P(f \mid w^1)] \\ &= E_{M_1}(q_1) = E_P[q_1(w^1)] = E_P(f). \qquad \text{Q.E.D.} \end{aligned}$$

We are given a Borel subset $Z \subset \mathcal{R}^m$, an open subset $\Phi \subset \mathcal{R}^w$, and a function $\psi\colon Z \times \Phi \to \mathcal{R}^a$ such that for each $\mu \in \Phi$, $\psi(\cdot, \mu)\colon Z \to \mathcal{R}^a$ is measurable, and for each

$z \in Z$, $\partial\psi/\partial\mu'$ is continuous on Φ. Let Q be a probability measure whose support is a subset of Z. We shall consider pairs (Q, μ) that satisfy the following condition:

CONDITION (C_1). (i) $\mu \in \Phi$; (ii) $E_Q[\psi(z,\mu)] = 0$; (iii) $E_Q[\psi(z,\mu)\psi'(z,\mu)]$ is positive-definite; (iv) $\operatorname{rank} E_Q[\partial\psi(z,\mu)/\partial\mu'] = w$.

Lemma 2 is taken from Chamberlain (1987, Lemma 1). (The continuity in z of ψ and $\partial\psi/\partial\mu'$ was assumed but not used in the proof.)

LEMMA 2. Suppose that (M, μ_0) satisfies Condition (C_1) and that M has finite support $\{\eta_1, \ldots, \eta_r\}$. Then there is an open set $A \subset \mathcal{R}^s$ ($s = r - (a - w) - 1$) and a family of probability measures $\{M_\gamma : \gamma \in A\}$ such that for all $\gamma \in A$: (i) M_γ has support $\{\eta_1, \ldots, \eta_r\}$; (ii) (M_γ, γ^1) satisfies Condition (C_1), where γ^1 contains the first w elements of γ; (iii) there is a $\gamma_0 \in A$ with $\gamma_0^1 = \mu_0$ and $M_{\gamma_0} = M$; (iv) $\pi_j(\gamma) \equiv M_\gamma\{\eta_j\}$ is a continuously differentiable function of γ ($j = 1, \ldots, r$), and

$$J_\gamma = \sum_{j=1}^{r} \pi_j^{-1}(\gamma) \frac{\partial \pi_j(\gamma)}{\partial \gamma} \frac{\partial \pi_j(\gamma)}{\partial \gamma'}$$

is positive definite.

PROOF OF THEOREM 2. Use Lemma 1 to construct M with finite support such that $(M, \theta_0, h_0) \in \Gamma$ and $V(M, \theta_0, h_0) = V_0$. With ψ and μ_0 defined as in the proof of Theorem 1, it follows that (M, μ_0) satisfies Condition (C_1). Hence we can use Lemma 2 to construct a family of probability measures $\{M_\gamma,\ \gamma \in A\}$ such that (M_γ, γ^1) satisfies Condition (C_1). Partition γ^1 into $\gamma^{1'} = (\theta', \lambda', \phi)$, where θ is $p \times 1$, λ is $l_T \times 1$, and ϕ is scalar. The support of M_γ coincides with that of M and so the M_γ-support of w^t is $\{\tau_{t1}, \ldots, \tau_{tl_t}\}$ (using the notation in the proof of Theorem 1). $E_{M_\gamma}\psi(z, \gamma^1) = 0$ implies that $E_{M_\gamma}[\rho_t(z, \theta, h(w^T)) \mid w^t = \tau_{tj}] = 0$ for $j = 1, \ldots, l_t$ provided that h is defined so that $h(\tau_{Tj}) = \lambda_j$ ($j = 1, \ldots, l_T$). Hence there is a $\delta > 0$ such that $(M_\gamma, \theta, h) \in \Gamma$ if $||\gamma - \gamma_0|| < \delta$. Let ζ be the s^{th} component of γ^1 if $s \leq p$; otherwise ζ is the final component. Then

$$\begin{aligned}
&\liminf_{n\to\infty} \sup_{(Q,\theta,h)\in\Gamma} E_Q\{l[\sqrt{n}(T_n - \alpha_s)]\} \\
&\quad \geq \liminf_{n\to\infty} \sup_{\|\gamma-\gamma_0\|<\delta} E_{M_\gamma}\{l[\sqrt{n}(T_n - \zeta)]\} \\
&\quad \geq (2\pi)^{-1/2} \int_{-\infty}^{\infty} l(\sigma_0 u) \exp(-\tfrac{1}{2}u^2)\, du,
\end{aligned}$$

where σ_0^2 is the (s,s) component of Ω^{-1} in (A.1) if $s \leq p$; otherwise σ_0^2 is the component in the lower-right corner. The second inequality follows from Theorem 1 in Chamberlain (1987), which shows that the information bound is given by Ω in the multinomial case. Then the proof of Theorem 1 shows that σ_0^2 is the (s,s) element of V_0. Q.E.D.

FOOTNOTES

* I am grateful to Ronald Gallant, Guido Imbens, Jinyong Hahn, Bo Honoré, and Whitney Newey for helpful discussions and comments. Financial support was provided by the National Science Foundation.

1 We shall generally omit the "with probability one" qualifier when we specify a conditional expectation.

2 Honoré (1990) extends this approach to include a lagged dependent variable.

3 We shall simplify the notation by omitting the P subscript from the expectation ($E = E_P$).

4 We shall simplify the notation by letting E denote E_P.

BIBLIOGRAPHY

Ahn, S. and P. Schmidt (1993): "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, forthcoming.

Andersen, E. B. (1970): "Asymptotic Properties of Conditional Maximum Likelihood Estimators," *Journal of the Royal Statistical Society*, Series B, 32, 283–301.

Anderson, T. W. and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47–82.

Arellano, M. and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

Arellano, M. and O. Bover (1993): "Another Look at the Instrumental-Variable Estimation of Error-Components Models," unpublished manuscript, CEMFI, Madrid.

Ashenfelter, O. (1978): "Estimating the Effect of Training Programs on Earnings," *Review of Economics and Statistics*, 60, 47–57.

Ashenfelter, O. and D. Card (1985): "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs," *Review of Economics and Statistics*, 67, 648–660.

Balestra, P. and M. Nerlove (1966): "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," *Econometrica*, 34, 585–612.

Bhargava, A. and J. D. Sargan (1983): "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods," *Econometrica*, 51, 1635–1659.

Chamberlain, G. (1984): "Panel Data," in Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics*, vol. 2, Amsterdam: North-Holland.

Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334.

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.

Hahn, J. (1991): "Efficient Estimation of Panel Data Models with Sequential Moment

Restrictions," unpublished manuscript, Harvard University.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Hansen, L. P., J. Heaton, and M. Ogaki (1988): "Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions," *Journal of the American Statistical Association*, 83, 863–871.

Hausman, J., B. Hall, and Z. Griliches (1984): "Econometric Models for Count Data with an Application to the Patents–R & D Relationship," *Econometrica*, 52, 909–938.

Hayashi, F. and C. Sims (1983): "Nearly Efficient Estimation of Time Series Models with Predetermined, But Not Exogenous, Instruments," *Econometrica*, 51, 783–798.

Holtz-Eakin, D., W. Newey, and H. Rosen (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371–1395.

Honoré, B. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60, 533–565.

Honoré, B. (1990): "Orthogonality Conditions for Tobit Models with Fixed Effects and Lagged Dependent Variables," unpublished manuscript, Northwestern University.

Keane, M. and D. Runkle (1992): "On the Estimation of Panel-Data Models with Serial Correlation When Instruments Are Not Strictly Exogenous," *Journal of Business and Economic Statistics*, 10, 1–9.

LaLonde, R. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 76, 604–620.

Manski, C. (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, 357–362.

Newey, W. (1992): "Efficient Estimation of Models with Conditional Moment Restrictions," unpublished manuscript, Massachusetts Institute of Technology; forthcoming in G. S. Maddala and C. R. Rao, eds., *Handbook of Statistics, Volume II: Econometrics*.

Newey, W. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.

Rasch, G. (1960): *Probabilistic Models for Some Intelligence and Attainment Tests*, Copenhagen: Denmarks Paedagogiske Institute.

Rasch, G. (1961): "On General Laws and the Meaning of Measurement in Psychology," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and*

Probability, Vol. 4, Berkeley: University of California Press.

Robinson, P. (1987): "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica*, 55, 875–891.

Robinson, P. (1991): "Best Nonlinear Three-Stage Least Squares Estimation of Certain Econometric Models," *Econometrica*, 59, 755–786.

Wooldridge, J. (1990): "Distribution-Free Estimation of Some Nonlinear Panel Data Models," unpublished manuscript, Massachusetts Institute of Technology.

Wooldridge, J. (1991): "Multiplicative Panel Data Models Without the Strict Exogeneity Assumption," unpublished manuscript, Massachusetts Institute of Technology.

Zeldes, S. (1989): "Consumption and Liquidity Constraints: An Empirical Investigation," *Journal of Political Economy*, 305–346.