:TRICS

:er Hill

ADVANCES IN ECONOMETRICS VOLUME 17

MAXIMUM LIKELIHOOD ESTIMATION OF MISSPECIFIED MODELS: TWENTY YEARS LATER

EDITED BY

THOMAS B. FOMBY

Department of Economics, Southern Methodist University, USA

R. CARTER HILL

Economics Department, Louisiana State University, USA

2003

ELSEVIER

JAI

Amsterdam – Boston – Heidelberg – London – New York – Oxford – Paris
San Diego – San Francisco – Singapore – Sydney – Tokyo

BAYESIAN ANALYSIS OF MISSPECIFIED MODELS WITH FIXED EFFECTS

Tiemen Woutersen

ABSTRACT

One way to control for the heterogeneity in panel data is to allow for time-invariant, individual specific parameters. This fixed effect approach introduces many parameters into the model which causes the "incidental parameter problem": the maximum likelihood estimator is in general inconsistent. Woutersen (2001) shows how to approximately separate the parameters of interest from the fixed effects using a reparametrization. He then shows how a Bayesian method gives a general solution to the incidental parameter for correctly specified models. This paper extends Woutersen (2001) to misspecified models. Following White (1982), we assume that the expectation of the score of the integrated likelihood is zero at the true values of the parameters. We then derive the conditions under which a Bayesian estimator converges at rate $\sqrt{N}$ where N is the number of individuals. Under these conditions, we show that the variance-covariance matrix of the Bayesian estimator has the form of White (1982). We illustrate our approach by the dynamic linear model with fixed effects and a duration model with fixed effects.

Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later
Advances in Econometrics, Volume 17, 235–249

ISSN: 0731-9053/doi:10.1016/S0731-9053(03)17011-9

1. INTRODUCTION

In applied work, economist rarely have data that can be viewed as being generated by an homogeneous group. That is, firms or individuals differ in observed and unobserved ways. These unobserved differences are usually referred to as heterogeneity and one can control for the heterogeneity in panel data by allowing for time-invariant, individual specific parameters. Accounting for heterogeneity using such individual or fixed effects avoids distributional and independence assumptions (which are usually not supported by economic theory), see Chamberlain (1984, 1985), Heckman et al. (1998) and Arellano and Honoré (2001).

This fixed effect approach introduces many parameters into the model which causes the "incidental parameter problem" of Neyman and Scott (1948): the maximum likelihood estimator is in general inconsistent. Chamberlain (1984), Trognon (2000) and Arellano and Honoré (2001) review panel data techniques that give good estimators for specific models. Woutersen (2001) derives a general solution that approximately separates the parameters of interest from the fixed effects using a reparametrization. After the reparametrization, the fixed effects are integrated out with respect to a flat prior. This yields a Bayesian estimator for the parameter of interest, $\hat{\beta}$, that has a low bias, $O(T^{-2})$ where T is the number of observations per individual. Moreover, the asymptotic distribution of $\hat{\beta}$ has the following form,

$$\sqrt{NT}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, I(\beta)^{-1}).$$

where $I(\beta)$ is the information matrix and $T \propto N^{\alpha}$ where $\alpha > 1/3$. Thus, the asymptotic variance of $\hat{\beta}$ is the same as the asymptotic variance of the infeasible maximum likelihood estimator that uses the true values of the fixed effects.

This paper extends the analysis of Woutersen (2001) by allowing for misspecification of the likelihood. Following White (1982), we assume that the expectation of the score is zero at the true values of the parameters. We then derive the primitive conditions under which the Bayesian estimator converges at rate $\sqrt{N}$. In particular, we assume the "score" of the integrated likelihood to be zero at the true value of the parameter of interest. Under these conditions, we show that the variance-covariance matrix of the Bayesian estimator has the form of White (1982).

Lancaster (2000, 2002) does not derive asymptotic variances and another new feature of this paper is that it derives the asymptotic variance of the integrated likelihood in a fixed T, increasing N asymptotics. We illustrate our approach by the dynamic linear model with fixed effects and a duration model with fixed effects.

This paper is organized as follows. Section 2 reviews information-orthogonality as a way to separate the nuisance parameters from the parameter of interest. Section 3 discusses the integrated likelihood approach. Section 4 gives the conditions for consistency and derives the variance-covariance matrix under misspecification. Section 5 discusses the dynamic linear model and a duration model and Section 6 concludes.

2. INFORMATION-ORTHOGONALITY

The presence of individual parameters in the likelihood can inhibit consistent estimation of the parameters of interest, as shown by Neyman and Scott (1948). For example, the dynamic linear model with fixed effects cannot be consistently estimated by maximum likelihood, as shown by Nickell (1981).[1] Information-orthogonality reduces the dependence between the parameters of interest and the individual parameters. We introduce more notation so that we can be specific. Suppose we observe N individuals for T periods. Let the log likelihood contribution of the tth spell of individual i be denoted by L^{it}. Summing over the contributions of individual i yields the log likelihood contribution,

$$L^{i}(\beta, \lambda_i) = \sum_t L^{it}(\beta, \lambda_i),$$

where β is the common parameter and λ_i is the individual specific effect. Suppose that the parameter β is of interest and that the fixed effect λ_i is a nuisance parameter that controls for heterogeneity. We can approximately separate β from $\lambda = \{\lambda_1, \ldots, \lambda_N\}$ by using an information-orthogonal parametrization of the quasi likelihood. In particular, information-orthogonality reduces this dependence between β and λ by having cross derivatives of the quasi log-likelihood being zero in expectation. That is,

$$EL_{\beta\lambda}(\beta_0, \lambda_0) = 0$$

i.e.

$$\int_{y_{\min}}^{y_{\max}} L_{\beta\lambda}(\beta_0, \lambda_0)\, e^{L(\beta_0, \lambda_0)}\, dy = 0,$$

where y denotes the dependent variable, $y \in [y_{\min}, y_{\max}]$ and $\{\beta_0, \lambda_0\}$ denote the true value of the parameters. Cox and Reid (1987) and Jeffrey (1961) use this concept and refer to it as "orthogonality." Lancaster (2000, 2002) applies this orthogonality idea to panel data and Woutersen (2000) gives an overview of orthogonality concepts.

Chamberlain (1984) and Arellano and Honoré (2001) review panel data econometrics in their handbook chapters. All but two of their models can be written in information-orthogonal form.[2]

Suppose that a quasi-likelihood is *not* information-orthogonal. In that case we reparameterize the quasi-likelihood to make it information-orthogonal. Let the individual nuisance parameter that is *not* information-orthogonal be denoted by f. We can interpret f as a function of β and information-orthogonal λ, $f(\beta,\lambda)$, and write the log likelihood as $L(\beta, f(\beta,\lambda))$. Differentiating $L(\beta, f(\beta,\lambda))$ with respect to β and λ yields

$$\frac{\partial L(\beta, f(\beta,\lambda))}{\partial\beta} = L_\beta + L_f \frac{\partial f}{\partial\beta}$$
$$\frac{\partial^2 L(\beta, f(\beta,\lambda))}{\partial\lambda\partial\beta} = L_{f\beta}\frac{\partial f}{\partial\lambda} + L_{ff}\frac{\partial f}{\partial\lambda}\frac{\partial f}{\partial\beta} + L_f\frac{\partial^2 f}{\partial\lambda\partial\beta}$$

where L_f is a score and therefore $EL_f = 0$. Information-orthogonality requires the cross-derivative $\partial^2 L(\beta, f(\beta,\lambda))/\partial\lambda\partial\beta$ to be zero in expectation, i.e.

$$EL_{\beta\lambda} = EL_{f\beta}\frac{\partial f}{\partial\lambda} + EL_{ff}\frac{\partial f}{\partial\lambda}\frac{\partial f}{\partial\beta} = 0.$$

This implies the following differential equation

$$EL_{f\beta} + EL_{ff}\frac{\partial f}{\partial\beta} = 0. \qquad (1)$$

If Eq. (1) has an analytical solution then $f(\cdot)$ can be written as a function of $\{\beta,\lambda\}$. If Eq. (1) has an implicit solution, then the Jacobian $\partial\lambda/\partial f$ can be recovered from the implicit solution. The Jacobian $\partial\lambda/\partial f$ is all we need for a reparametrization in a Bayesian framework. The general nonlinear model and the single index model have an information-orthogonal parametrization that is implicit, as shown in Woutersen (2001). For the remainder of the paper, we assume information-orthogonality.

The "invariance result" of the maximum likelihood estimator implies that reparametrizations do not change the estimates. In particular, an information-orthogonal parametrization would yield the same estimates for β as a parametrization that is not information-orthogonal. However, the integrating out method does not have this invariance property and this paper shows that information-orthogonality can yield moment functions that are robust against incidental parameters, even under misspecification.

3. THE INTEGRATED LIKELIHOOD

After ensuring information-orthogonality, we integrate out the fixed effects and use the mode of the integrated likelihood as an estimator. That is,

$$\hat\beta = \arg\max_\beta L^I(\beta)$$

where

$$L^I(\beta) = \sum_i \ln\left(\int e^{L^i(\beta,\lambda)} d\lambda_i\right).$$

Misspecification has been, so far, not considered in combination with the integrated likelihood approach as is apparent from the overviews of Gelman et al. (1995) and Berger et al. (1999). The point of this paper, however, is to consider misspecification. In particular, $L^i(\beta,\lambda)$ does not need to be a fully specified likelihood. It is sufficient that we specify, as an approximation, a density for y_{it} that is conditional on x_{it} and λ_i. The likelihood contribution $L^{it}(\beta,\lambda)$ is the logarithm of this conditional density and $L^i(\beta,\lambda) = \sum_t L^{it}(\beta,\lambda)$. In particular, the distribution of the fixed effects is left unrestricted. Thus, in this set-up we can think of the Data Generating Process as follows. First, the fixed effects, $f_1,\ldots,f_N$, are generated from an unknown and unrestricted distribution. As a second step, $x_{11},\ldots,x_{N1}$ is generated from another unknown distribution that can depend on $f_1,\ldots,f_N$. Then $y_{11},\ldots,y_{N1}$ is generated by a conditional distribution[3] that is approximated by the econometrician. For period $t = 2$, the distribution of $x_{12},\ldots,x_{N2}$ can depend on $f_1,\ldots,f_N, x_1,\ldots,x_N$. Alternatively, x_{it} can allowed to be endogenous in which case the econometrician specifies a density for y_{it} that is conditional on $x_{i,t-1}$ and f_i.

4. ASSUMPTIONS AND THEOREM

In this section, we consider estimation while allowing for misspecification of the model. The clearest approach seems to impose the assumptions directly on the integrated likelihood function. White (1982, 1993) assumes that the expectation of the score is zero at the true value of the parameter. Similarly, we assume that the score of the *integrated* likelihood has expectation zero at the truth.

Assumption 1. (i) Let $\{x_i, y_i\}$ be i.i.d. and (ii) let $EL_\beta^{i,I} = 0$ for every i.

This assumption implies, by independence across individuals, that $EL_\beta^{i,I} L_\beta^{j,I} = 0$ for $i \neq j$. Note that the regressor $x_i = \{x_{i1}, x_{i2}, \ldots, x_{iT}\}$ and dependent variable

$y_i = \{y_{i1}, y_{i2}, \ldots, y_{iT}\}$ are not required to be stationary and that x_i is not required to be exogenous.

Assumption 2. (i) $\beta \in \Theta$ where which is compact or (ii) $L^I_\beta(\beta)' L^I_\beta(\beta)$ is concave in β.

This is a regularity condition that is often assumed.

Assumption 3. (i) $EL^{\prime I}_\beta(\beta) = 0$ is uniquely solved for $\beta = \beta_0$; (ii) $L^{\prime I}_\beta(\beta)$ is continuous at each $\beta \in \Theta$ with probability one; and (iii) $E \sup_{\beta\in\Theta} ||L^I_\beta(\beta)|| < \infty$.

Information-orthogonality, $EL_{\beta\lambda}(\beta_0, \lambda_0) = 0$, does not imply $EL^{\prime I}_\beta(\beta) = 0$ but the stronger condition $L_{\beta\lambda}(\beta, \lambda) = 0$ does. However, imposing this stronger condition excludes many interesting models. Thus, it could be that $EL_{\beta\lambda}(\beta_0, \lambda_0) = 0$ is not a necessary condition for $EL^I_\beta(\beta) = 0$ but we do not know examples for which $EL^{\prime I}_\beta(\beta) = 0$ and $EL_{\beta\lambda}(\beta_0, \lambda_0) \neq 0$. We therefore recommend to first reparameterize the model so that $EL_{\beta\lambda}(\beta_0, \lambda_0) = 0$ and, as a second step, check Assumptions 1–3.

Assumption 4. (i) $\beta_0 \in$ interior(Θ); (ii) $L^I_\beta(\beta)$ is continuously differentiable in a neighborhood $\mathfrak{N}$ of β_0; (iii) $EL^I_{\beta\beta}(\beta)$ is continuous at β_0 and $\sup_{\beta\in\mathfrak{N}} ||EL^I_{\beta\beta}(\beta) - EL^I_{\beta\beta}(\beta)|| \xrightarrow{p} 0$; and (iv) $EL^I_{\beta\beta}(\beta_0)$ is nonsingular.

Theorem 1. Suppose $\hat\beta = \arg\min_\beta\{(L^I_\beta(\beta)/NT')(L^I_\beta(\beta)/NT)\}$. Let Assumptions 1–4 hold. Let $N \to \infty$ while T is fixed. Then

$$\sqrt{NT}(\hat\beta - \beta_0) \to N(0, \Psi)$$

where

$$\Psi = \left[\frac{1}{NT} EL^I_{\beta\beta}(\beta_0)\right]^{-1} \left[\frac{1}{NT} E\{(L^I_\beta(\beta_0))(L^I_\beta(\beta_0))'\}\right] \left[\frac{1}{NT} EL^I_{\beta\beta}(\beta_0)\right]^{-1}.$$

Proof: See Appendix A. □

The theorem shows that the integrated likelihood as a convenient tool to derive moments that are robust against incidental parameters as well as robust against misspecification of the parametric error term.

5. EXAMPLES

In this section we discuss two examples that illustrate the integrated likelihood approach.

5.1. *Dynamic Linear Model*

Consider the dynamic linear model with fixed effects,

$$y_{it} = y_{i,t-1}\beta + f_i + \varepsilon_{it} \quad \text{where} \quad E\varepsilon_{it} = 0, E\varepsilon^2_{it} < \infty$$
$$\text{for} \quad E\varepsilon_{is}\varepsilon_{it} = 0 \quad \text{for} \quad s \neq t \quad \text{and} \quad t = 1, \ldots, T.$$

This model is perhaps the simplest model that nests both state dependence and heterogeneity as alternative explanations for the variation in the values of y_{it} across agents. As such, the dynamic linear model is popular in the development and growth literature. For a discussion and further motivation of this model, see Kiviet (1995), Hahn, Hausman and Kuersteiner (2001), Arellano and Honoré (2001) as well as the references therein. Lancaster (2002) suggests the following information-orthogonal parametrization,

$$f_i = y_{i0}(1 - \beta) + \lambda_i \,e^{-b(\beta)} \quad \text{where} \quad b(\beta) = \frac{1}{T}\sum_{t=1}^{T} \frac{T-t}{t}\beta^t.$$

However, Lancaster (2002) does not derive the asymptotic variance of the integrated likelihood estimator. Woutersen (2001) shows that, under normality of ε_{it}, the integrated likelihood estimator is adaptive for an asymptotic with $T \propto N^\alpha$ and $\alpha > 1/3$. That is, the asymptotic variance does not depend on knowledge of λ in this asymptotic. We now consider the case where the normality of ε_{it} fails to hold and only assume normality in order to derive the integrated likelihood estimator. Note that

$$EL^i_\beta = \frac{1}{\sigma^2} E \sum_t \varepsilon_t \{y_{t-1} - y_0 - b'(\beta)\lambda \,e^{-b(\beta)}\} = 0$$

$$EL^i_{\beta\lambda} = -\frac{b'(\beta)\,e^{-b(\beta)}}{\sigma^2} E \sum_t \varepsilon_t = 0.$$

The log likelihood contribution has the following form,

$$L^i = -\frac{1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_t (\tilde y_t - \tilde y_{t-1}\beta - \lambda \,e^{-b(\beta)})^2 \quad \text{where} \quad \tilde y_t = y_t - y_0.$$

Integrating with respect to λ gives the integrated likelihood contribution,

$$\begin{aligned} e^{L^{i,I}} &= \frac{1}{\sqrt{\sigma^2}} \int e^{-(1/2\sigma^2)\sum_t(\bar{y}_t - \bar{y}_{t-1}\beta - \lambda e^{-b(\beta)})^2}\, d\lambda \\ &= \frac{1}{\sqrt{\sigma^2}} e^{b(\beta)} \int e^{-(1/2\sigma^2)\sum_t(y_t - y_{t-1}\beta - f)^2}\, df \\ &= \frac{1}{\sqrt{\sigma^2}} e^{b(\beta)-(1/2\sigma^2)\sum_t(y_t - y_{t-1}\beta)^2} \int e^{-(T/2\sigma^2)\{f^2 - 2f\overline{(y_t - y_{t-1}\beta)}\}}\, df \\ &\propto e^{b(\beta)-(1/2\sigma^2)\{(T/2)\overline{(y_t - y_{t-1}\beta)}^2 + \sum_t (y_t - y_{t-1}\beta)^2\}}, \end{aligned}$$

where we omit the subscript i and $\partial\lambda/\partial f = e^{b(\beta)}$ does *not* depend on f. Taking logarithms and differentiating with respect to β yields

$$L_\beta^{i,I} = b'(\beta) + \frac{1}{\sigma^2}\sum_t (y_{it} - y_{i,t-1}\beta)y_{i,t-1} - T\overline{(y_{it} - y_{i,t-1}\beta)}\,\overline{y_{i,t-1}}$$

$$L_{\beta\beta}^{i,I} = b''(\beta) - \frac{1}{\sigma^2}\sum_t y_{i,t-1}^2 + \frac{1}{\sigma^2} T\overline{y_{i,t-1}}^2$$

where $b(\beta) = 1/T\sum_{t=1}^T (T - t/t)\beta^t$, $b(\beta)' = 1/T\sum_{t=1}^T (T-t)\beta^{t-1}$, $b(\beta)'' = 1/T\sum_{t=1}^T (T-t)(t-1)\beta^{t-2}$. Note that $EL_\beta^{i,I}/NT = 0$ for any N, T and that the mode of $L^I(\beta)/NT$ is a consistent estimator for β for N increasing. Analogue to the quasi-maximum likelihood estimator of White (1982), the asymptotic variance has the form of Theorem 1, $\Psi = [(1/NT)EL_{\beta\beta}^I]^{-1}[(1/NT)E\{(L_\beta^I)(L_\beta^I)'\}][(1/NT)EL_{\beta\beta}^I]^{-1}$. The author views the integrated likelihood as a convenient way to derive moments that can be robust against misspecification of the parametric error term. In particular, the parametric assumptions on the error term are irrelevant for the models with additive error terms that are discussed in Arellano and Honoré (2001).

5.2. *Duration Model with Time-Varying Individual Effects*

Consider a duration model in which the hazard depends on an individual effect f_i, a spell-specific effect u_{is} and observable regressors x_{is}. In particular, consider the following hazard,

$$\theta_{is}(t) = e^{f_i + x_{is}\beta + u_{is}}. \qquad (2)$$

where the subscript i refers an individual and the subscript s refers to a spell of that individual. This hazard depends on two unobservable stochasts, f_i and u_{is}.

In particular, the individual specific effect f_i can depend on the regressors x_{is}. We avoid distributional assumption on the spell-specific effect u_{is} but we assume that u_{is} is independent of x_{is} and $Ee^{-u_{is}} < \infty$. Thus, the hazard of Eq. (2) is a generalization of the fixed effect hazard model with regressors where the hazard is $e^{f_i + x_{is}\beta}$. Chamberlain (1984) developed an estimator for the last model and Van den Berg (2001) gives a current review of duration models. A common criticism of the model with hazard $e^{f_i + x_{is}\beta}$ is that it assumes that variations in the hazard can all be explained by variations in the regressor x_{is}. In other words, the unobservable effect is constant over time, see Van den Berg (2001) for this argument. Equation (2) extends this model by allowing for a spell-specific effect u_{it}. As an approximation of the model of (2) we consider

$$\theta_{is} = e^{\lambda_i + x_{is}\beta}$$

where $\sum_s x_{is} = 0$. This hazard implies a log likelihood and the normalization, $\sum_s x_{is} = 0$, ensures that the log likelihood is information-orthogonal. In particular,

$$L^i(\beta, \lambda_i) = T\lambda_i - e^{\lambda_i}\sum_s e^{x_{is}\beta}t_{is},$$

$$L_\beta^i(\beta, \lambda_i) = -e^{\lambda_i}\sum_s x_{is}\, e^{x_{is}\beta}t_{is},$$

and

$$L_{\beta\lambda_i}^i(\beta, \lambda_i) = -e^{\lambda_i}\sum_s x_{is}\, e^{x_{is}\beta}t_{is}.$$

Note that $e^{x_{is}\beta_0}t_{is}$ is exponentially distributed with mean $e^{-(\lambda_{0,i}+u_{is})}$. This implies,

$$EL_\beta^i(\beta_0, \lambda_{i,0}) = -E\, e^{\lambda_{0,i}}\sum_s x_{is}\, e^{-(\lambda_{0,i}+u_{is})} = -E\sum_s x_{is}\, e^{-u_{is}} = 0$$

since $\sum_s x_{is} = 0$ and u_{is} is independent of x_{is}. Similarly, $EL_{\beta\lambda_i}^i(\beta_0, \lambda_{i,0}) = 0$. Integrating λ_i with respect to the likelihood gives

$$L^{i,I} = \ln\int e^{L^i}\, d\lambda_i = \ln\int e_i^{T\lambda_i} e^{-\sum_s e^{x_{is}\beta + \lambda_i}t_{is}}\, d\lambda_i = \ln\left[\frac{\Gamma(T)}{\{\sum_s e^{x_{is}\beta}t_{is}\}^T}\right].$$

see Appendix B for details. Thus,

$$\frac{L_\beta^{i,I}}{T} = \frac{\sum_s x_{is}\, e^{x_{is}\beta}t_{is}}{\sum_s e^{x_{is}\beta}t_{is}}$$

and

$$\frac{L^I_\beta}{NT} = \frac{\sum_i}{N} \frac{\sum_s x_{is}\, e^{x_{is}\beta} t_{is}}{\sum_s e^{x_{is}\beta} t_{is}}$$

$$\frac{L^I_{\beta\beta}}{NT} = \frac{\sum_i}{N} \frac{\sum_s x^2_{is}\, e^{x_{is}\beta} t_{is} - \{\sum_s x_{is}\, e^{x_{is}\beta} t_{is}\}^2}{\{\sum_s e^{x_{is}\beta} t_{is}\}^2}$$

In Appendix C, it is shown that $(1/NT)EL^I_\beta = 0$ for any N and any $T \geq 2$. Thus, the mode of $L^I(\beta)/NT$ is a consistent estimator for β for N increasing. Moreover, the asymptotic variance has the form of Theorem 1, $\Psi = [(1/NT)EL^I_{\beta\beta}]^{-1}[(1/NT)E\{(L^I_\beta)(L^I_\beta)'\}][(1/NT)EL^I_{\beta\beta}]^{-1}$.

5.2.1. Simulation

Let the data be generated by the following hazard model,

$$\theta_{is}(t) = e^{f_i + x_{is}\beta + u_{is}}. \tag{3}$$

This hazard implies that the expected duration, conditional on f_i, x_{is}, and u_{is} equals[4] $1/e^{f_i + x_{is}\beta + u_{is}}$, i.e. $E(t_{is}|f_i, x_{is}, u_{is}) = 1/e^{f_i + x_{is}\beta + u_{is}}$. Let the exponent of the individual effect, e^{f_i}, have a unit exponential distribution and let the individual spell effect, u_{is}, be normally distributed with mean zero and variance σ^2_u. Suppose that we observe a group of N individuals and that we observe an unemployment spell before and after treatment, that is $x_{i1} = 0$ for all i and $x_{i2} = 1$ for all i. Heckman, Ichimura, Smith and Todd (1998) discuss the estimation of treatment effect models and conclude that the fixed effect model performs very well. This simulation study extends the fixed effect duration model by allowing for an spell specific effect u_{is}, $i = 1, \ldots, N$ and $s = 1, 2$. In particular, the model of Eq. (3) also extends both Chamberlain (1985) and Ridder and Woutersen (2003) by allowing for both random and fixed effects. We first assume that the treatment has no effect on the hazard out of unemployment, that is, $\beta = 0$. We then assume that the hazard out of unemployment increases by factor 2.7, That is, $\beta = 1$ and $e^\beta = e \approx 2.7$. The estimator developed in this subsection is denoted by "integrated likelihood estimator." A naive Bayes estimator that just integrated out the fixed effects and then uses the posterior mode is denoted by "naive Bayes estimator." We use flat priors for all parameters and base inference on the posterior mode after integrating out the fixed effects f_i, $i = 1, \ldots, N$. The model is misspecified in the sense that the individual spell effect, u_{is}, is ignored.

	Bias ($\beta = 0$)	RMSE ($\beta = 0$)	Bias ($\beta = 1$)	RMSE ($\beta = 1$)
Integrated likelihood estimator				
$\sigma^2_u = \frac{1}{2}$	−0.0008	0.1334	0.0039	0.1298
$\sigma^2_u = 1$	0.0145	0.1451	−0.0010	0.1467
$\sigma^2_u = 2$	−0.0008	0.1790	−0.0042	0.1826
Naive bayes estimator				
$\sigma^2_u = \frac{1}{2}$	1.1346	1.1424	1.1308	1.1394
$\sigma^2_u = 1$	1.2197	1.2288	1.2188	1.2285
$\sigma^2_u = 2$	1.3739	1.3856	1.3674	1.3795

Note that the two estimators use the same likelihood and priors. However, the "info-ortho Bayes estimator" separates the nuisance parameter from the parameter of interest before integrating out f_i, $i = 1, \ldots, N$. As a consequence, the bias is much lower, by about factor 8, for the "integrated likelihood estimator." Note that, for both estimators, the Root Mean Squared Error (RMSE) is increasing in σ^2_u and that the bias of the "naive Bayes estimator" does not strongly depend on the value of β. We conclude that separating the nuisance parameter from the parameter of interest works well for this misspecified model.

6. CONCLUSION

This paper extends the integrated likelihood estimator to misspecified models. Using information-orthogonality, we approximately separate the nuisance parameter from the parameter of interest. We use a Bayesian techniques since reparametrization of a nuisance parameter only requires an expression of the Jacobian in a Bayesian framework. Under the condition that the score of the integrated likelihood has expectation zero at the truth, we show that the variance-covariance matrix of the Bayesian estimator has the form of White (1982). Thus, information-orthogonality combined with the integrated likelihood is a promising approach which solves the incidental parameter problem of Neyman and Scott (1948) for a class of misspecified models. We illustrate our approach by two misspecified models with individual effects. In the dynamic linear model, we allow the error term to be non-normal and in the hazard model we allow the individual effect to change over time.

NOTES

1. The dynamic linear model assumes that $y_{it} = y_{i,t-1}\beta + f_i + \varepsilon_{it}$ and we discuss this model in Section 5.1.

2. The transformation model of Abrevaya (1998) and one discrete choice model by Honor'e and Kyriazidou (2000) are not information-orthogonal. Both models require infinite support for the regressor, can be estimated using a sign function and will be discussed in a separate paper that deals with "information-orthogonality" of sign functions.

3. That is, conditional on $f_1, \ldots, f_N$ and $x_1, \ldots, x_N$.

4. Note that t_{is} is exponentially distributed if we condition on f_i, x_{is}, and u_{is}.

ACKNOWLEDGMENTS

I gratefully acknowlegde stimulating suggestions from Tony Lancaster. The Social Science and Humanities Research Council of Canada provided financial support. All errors are mine.

REFERENCES

Abrevaya, J. (1998). *Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable*. Unpublished manuscript, Graduate School of Business, University of Chicago.

Arellano, M., & Honoré, B. E. (2001). Panel data models: Some recent developments. In: J. Heckman & E. Leamer (Eds), *Handbook of Econometrics* (Vol. 5). Amsterdam: North-Holland.

Berger, J. O., Liseo, B., & Wolpert, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, *14*, 1–28.

Chamberlain, G. (1984). Panel data. In: Z. Griliches & M. D. Intriligator (Eds), *Handbook of Econometrics* (Vol. 2). Amsterdam: North-Holland.

Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In: J. J. Heckman & B. Singer (Eds), *Longitudinal Analysis of Labor Market Data*. Cambridge: Cambridge University Press.

Cox, D. R., & Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society, Series B*, *49*, 1–39.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. New York: Chapman & Hall.

Hahn, J., Hausman, J., & Kuersteiner, G. (2001). Bias corrected instrumental variables estimation for dynamic panel models with fixed effects. MIT Working Paper.

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica*, *66*, 1017–1098.

Honoré, B. E., & Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, *68*, 839–874.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Clarendon Press.

Kiviet (1995). On bias, inconsistency and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics*, *68*, 53–78.

Lancaster, T. (2000). The incidental parameters since 1948. *Journal of Econometrics*, *95*, 391–413.

Lancaster, T. (2002). Orthogonal parameters and panel data. *Review of Economic Studies* (forthcoming).

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In: R. F. Engle & D. MacFadden (Eds), *Handbook of Econometrics* (Vol. 4). Amsterdam: North-Holland.

Neyman, J., & Scott, E. L. (1948). Consistent estimation from partially consistent observations. *Econometrica*, *16*, 1–32.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, *49*, 1417–1426.

Ridder, G., & Woutersen, T. M. (2003). The singularity of the efficiency bound of the mixed proportional hazard model. *Econometrica* (forthcoming).

Trognon, A. (2000). Panel data econometrics: A successful past and a promising future. Working Paper, Genes (INSEE).

Van den Berg, G. T. (2001). Duration models: Specification, identification, and multiple duration. In: J. Heckman & E. Leamer (Eds), *Handbook of Econometrics* (Vol. 5). Amsterdam: North-Holland.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*, 1–25.

White, H. (1993). *Estimation, inference and specification analysis*. New York: Cambridge University Press.

Woutersen, T. M. (2000). Consistent estimation and orthogonality. Working Paper, Department of Economics, University of Western Ontario.

Woutersen, T. M. (2001). Robustness against incidental parameters and mixing distributions. Working Paper, Department of Economics, University of Western Ontario.

APPENDIX A
THEOREM 1

To be shown

$$\sqrt{N}(\hat{\beta} - \beta_0) \rightarrow N(0, \Psi)$$

where

$$\Psi = \left[\frac{1}{NT} EL^{I}_{\beta\beta}(\beta_0)\right]^{-1} \left[\frac{1}{NT} E\{(L^{I}_{\beta}(\beta_0))(L^{I}_{\beta}(\beta_0))'\}\right] \left[\frac{1}{NT} EL^{I}_{\beta\beta}(\beta_0)\right]^{-1}.$$

Proof:

Let Assumptions 1, 2(i), and 3 hold. Then all the conditions of Newey and McFadden (1994, Theorem 2.6) are satisfied and consistency follows. Assuming that, in addition Assumption 4 holds then the assumptions of Newey and McFadden (1994, Theorem 3.2) are satisfied and asymptotic normality follows where the identity matrix is used a the weighting matrix.

Instead of assuming that the parameter space is compact as in 2(ii) we can assume that we assume that β_0 is an element of the interior of a convex set Θ and

$L^I_\beta(\beta)$ is concave for all i as in Assumption 2(ii) and 4(i). All the requirements of Newey and McFadden (1994, Theorem 2.7) are satisfied and consistency of the integrated likelihood estimator follows. Asymptotic normality is again implied by Newey and McFadden (1994, Theorem 3.2).

APPENDIX B
DURATION EXAMPLE, INTEGRATED LIKELIHOOD

To be shown,

$$L^{i,I} = \ln \int e_i^{T\lambda_i} e^{-\sum_s e^{x_{is}\beta+\lambda_i} t_{is}} d\lambda_i = \ln \left[\frac{\Gamma(T)}{\{\sum_s e^{x_{is}\beta} t_{is}\}^T} \right]$$

where $\Gamma(\cdot)$ denotes the Gamma function

Proof: Define $v_i = e^{\lambda_i}$.

$$L^{i,I} = \ln \int e^{L^i} \frac{1}{v_i} dv_i = \ln \int v_i^{T-1} e^{-v_i \sum_s e^{x_{is}\beta} t_{is}} dv_i$$

$$= \ln \left\{ \frac{\Gamma(T)}{\{\sum_s e^{x_{is}\beta} t_{is}\}^T} \int \frac{\{\sum_s e^{x_{is}\beta} t_{is}\}^T v_i^{T-1} e^{-v_i \sum_s e^{x_{is}\beta} t_{is}}}{\Gamma(T)} dv_i \right\}.$$

Note that $(\{\sum_s e^{x_{is}\beta} t_{is}\}^T v_i^{T-1} e^{-v_i \sum_s e^{x_{is}\beta} t_{is}})/\Gamma(T)$ is a gamma density with parameters T and $\sum_s e^{x_{is}\beta} t_{is}$ and that this density integrates to one. The result follows. Q.E.D.

APPENDIX C
DURATION EXAMPLE, SCORE

To be shown,

$$\frac{1}{NT} EL^I_\beta = 0 \quad \text{where} \quad \frac{L^I_\beta}{NT} = \frac{\sum_i}{N} \frac{\sum_s x_{is} e^{x_{is}\beta_0} t_{is}}{\sum_s e^{x_{is}\beta_0} t_{is}}$$

for any N and any $T \geq 2$.

Proof:

$$\frac{1}{NT} EL^I_\beta = E \frac{\sum_i}{N} \frac{\sum_s x_{is} e^{x_{is}\beta_0} t_{is}}{\sum_s e^{x_{is}\beta_0} t_{is}} = E \frac{\sum_i}{N} \frac{\sum_s x_{is} e^{x_{is}\beta_0+\lambda_0} t_{is}}{\sum_s e^{x_{is}\beta_0+\lambda_0} t_{is}}$$

Note that $e^{x_{is}\beta_0+\lambda_0} t_{is}$ is exponentially distributed with mean $e^{-u_{is}}$. Also note that the expectation of $e^{x_{is}\beta_0+\lambda_0} t_{is} / \sum_s e^{x_{is}\beta_0+\lambda_0} t_{is}$ does *not* depend on x_{is}. Thus, define $\mu_i = E(e^{x_{is}\beta_0+\lambda_0} t_{is}) / \sum_s e^{x_{is}\beta_0+\lambda_0} t_{is}$. This yields

$$\frac{1}{NT} EL^I_\beta = E \frac{\sum_i}{N} \sum_s x_{is} = E \frac{\sum_i}{N} \mu_i \sum_s x_{is} = 0$$

since $\sum_s x_{is} = 0$. Q.E.D.