PROFILE-SCORE ADJUSTMENTS FOR NONLINEAR FIXED-EFFECT MODELS

Geert Dhaene* Koen Jochmans†
K.U. Leuven CORE

March 15, 2011

PRELIMINARY AND INCOMPLETE

Abstract: Maximum-likelihood estimates of nonlinear panel data models with fixed effects are generally not consistent as the number of units, N, grows large while the number of time periods, T, stays fixed. The inconsistency can be viewed as a consequence of the bias of the score function, where the unit-specific parameters have been profiled out. We investigate ways of adjusting the profile score so as to make it unbiased or approximately unbiased. This leads to estimators, solving an adjusted profile score equation, that are fixed-T consistent or have less asymptotic bias, as $T \to \infty$, than maximum likelihood. One approach to adjusting the profile score is to subtract its bias, evaluated at maximum-likelihood estimates of the fixed effects. When this bias does not depend on the incidental parameters, the adjustment is exact. Otherwise, it does not eliminate the bias entirely but reduces its order (in T), and it can be iterated, reducing the bias order further. We examine a range of nonlinear models with additive fixed effects. In many of these, an exact bias adjustment of the profile score is possible. In others, suitably adjusted profile scores exhibit much less bias than without the adjustment, even for very small T.

JEL classification: C13, C15, C23, C25
Keywords: adjusted profile score, bias correction, fixed effects, nonlinear models.

Introduction

Consider the problem of inferring the value of a finite-dimensional parameter θ in a parametric model from a panel data set consisting of T observations on N units. In microeconometric models, unit-specific parameters, called fixed effects or incidental parameters, are often included to account for unobserved heterogeneity. For example, in the agricultural production-function application of Mundlak (1961), firm-specific intercepts serve to control for the impact of managerial ability and soil quality on firm output. Alternatively, Hausman, Hall, and Griliches (1984) and Hospido (2010) introduced fixed effects to allow for heterogeneity in dispersion parameters in applications to the patents–R&D relationship and the volatility of wages, respectively. Unfortunately, including fixed effects generally renders the maximum-likelihood estimator (MLE)

*Address: K.U. Leuven, Department of Economics, Naamsestraat 69, B-3000 Leuven, Belgium. Tel. +32 16 326798; Fax +32 16 326796; E-mail: `geert.dhaene@econ.kuleuven.be`.

†Address: U.C. Louvain, Center for Operations Research and Econometrics, Voie du Roman Pays 34, B-1348 Louvain-la-Neuve, Belgium. Tel. +32 10 474329; E-mail: `koen.jochmans@uclouvain.be`.

of θ inconsistent if T remains fixed while $N \to \infty$ (Neyman and Scott, 1948). The problem is known as the incidental-parameter problem.

The profile score function replaces the fixed effects with their maximum-likelihood estimates for a given θ and, therefore, is a feasible version of the score function that would be used if the fixed effects were known. This replacement generally induces a bias of order $O(T^{-1})$. The MLE, $\widehat{\theta}$, sets the profile score to zero and therefore inherits this bias. There are important situations where alternative estimating equations are available that are free of fixed effects (see, e.g., Arellano and Honoré, 2001, for an overview). However, there is no general method for deriving such estimating equations and they may not exist simply because θ may not be fixed-T point-identified (see Chamberlain, 2010). In this paper we seek to adjust the profile score, following McCullagh and Tibshirani (1990). The key element is a calculation of the bias of the profile score, either analytically or via simulation, which is then evaluated at maximum-likelihood estimates of the fixed effects. If the bias is free of fixed effects, this leads to an unbiased estimating equation. Otherwise, it results in an estimating equation whose bias is $O(T^{-2})$. We show that it is possible to iterate the adjustment, yielding adjusted profile scores with bias of successively smaller order, $O(T^{-2}), O(T^{-3}), \ldots$ Depending on the situation at hand, the adjustments give rise to estimators that are either fixed-T consistent or have a smaller order of bias than the MLE. Our approach fits into the literature on bias-corrected fixed-effect estimation recently surveyed by Arellano and Hahn (2007) and inference from integrated likelihoods (Lancaster, 2002, and Arellano and Bonhomme, 2009), and the parallel developments in the statistics literature (e.g., Li, Lindsay, and Waterman, 2003, and Sartori, 2003).

Focusing on the profile score rather than on $\widehat{\theta}$ directly has some advantages. First, it offers a direct way of verifying whether the presence of incidental parameters effectively leads to the inconsistency of the MLE. For example, in the fixed-effect Poisson and exponential-regression models a short calculation suffices to show that the MLE is consistent. On the other hand, verifying whether θ and the incidental parameters are likelihood orthogonal, which is sufficient for the consistency of the MLE, may be a cumbersome task, especially because that may be true in one parametrization but not in another (see Lancaster, 2000). For the Poisson model, for instance, the equivalence between maximum likelihood and the conditional-likelihood estimator introduced in Hausman, Hall, and Griliches (1984) was not known until Lancaster (2002) and Blundell, Griffith, and Windmeijer (2002). Second, there are several models of practical interest where the bias of the profile score, although non-zero, is free of incidental parameters while the magnitude of $\text{plim}_{N\to\infty}\widehat{\theta}$ depends on the distribution of incidental parameters and covariates. One important model where this is the case is the linear dynamic fixed-effect model (Nickell, 1981; Dhaene and Jochmans, 2010b). Weibull and gamma duration models are other examples; details are provided below. In such cases, unbiased estimating equations and fixed-T consistent estimators can be formed by centering of the profile score, a point already made by

Neyman and Scott (1948). Consequently, adjusting the profile score can be much simpler than approaches targeted at adjusting the MLE directly; see the approaches described in MacKinnon and Smith (1998), for example. In addition, there is an asymptotic justification, complementing the discussion and motivations offered in McCullagh and Tibshirani (1990), for using an adjusted profile score also in situations where the expected profile score does depend on the fixed effects. Fourth, the adjustments, including the iterated adjustment, are easy to carry out. They do not require explicit knowledge of the dependence of the bias on the fixed effects. This is in contrast to the bias correction methods in Arellano and Hahn (2006), Carro (2007), and Bester and Hansen (2009). Fifth, the iterative procedure leads to higher-order bias adjustments, as does the jackknife (Dhaene and Jochmans, 2010a).

Section 2 presents the profile score adjustment and how it can be iterated. We discuss examples in Section 3, mostly to nonlinear models. Section 3 illustrates the gains of the adjustments by simulations in the context of static and dynamic binary-choice models.

1 Adjusting the profile score

We are given a panel data set (y_{it}, x_{it}) where $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Assume independence across i. The conditional density of y_{it} given x_{it}, $f(y_{it}|x_{it}; \theta, \eta_i)$, is known up to the common parameter θ and the unit-specific parameters η_i. Both θ and η_i may be vectors. We are interested in estimating θ. Since Neyman and Scott (1948), it is known that the maximum-likelihood estimate (MLE) of θ need not be consistent as $N \to \infty$ with T fixed. One may view the inconsistency as resulting from a biased profile score function. The (normalized) profile log-likelihood and score functions, and their ith contributions, are

$$\begin{aligned} l &= \frac{1}{N}\sum_i l_i, & l_i = l_i(\theta) = \frac{1}{T}\sum_t \log f(y_{it}|x_{it}; \theta, \widehat{\eta}_i), \\ s &= \frac{1}{N}\sum_i s_i, & s_i = s_i(\theta) = \frac{1}{T}\sum_t \nabla_\theta \log f(y_{it}|x_{it}; \theta, \widehat{\eta}_i), \end{aligned}$$

where $\widehat{\eta}_i$ is the MLE of η_i for a given θ,

$$\widehat{\eta}_i = \widehat{\eta}_i\,(\theta) = \arg\max_{\eta_i} \frac{1}{T}\sum_t \log f(y_{it}|x_{it}; \theta, \eta_i).$$

Assuming that f is sufficiently regular, the MLE solves $s = 0$ for θ. Let $E = E_{\theta,\eta_i}$ denote the expectation operator at true parameter values, with exogenous variates (and, possibly, initial observations) held fixed at observed values. As is well known, the expected score vanishes at the true value, i.e. $E\frac{1}{T}\sum_t \nabla_\theta \log f(y_{it}|x_{it}; \theta, \eta_i) = 0$. However, the profile score replaces η_i with $\widehat{\eta}_i$. Except in special cases, this makes Es_i and its aggregate Es nonzero, causing $s = 0$ to be a biased estimating equation and the MLE to be inconsistent. Under regularity conditions,

$Es_i = O(T^{-1}) = Es$ as $T \to \infty$ and the $O(T^{-1})$ bias of the profile score carries over to the MLE, that is, the asymptotic bias of the MLE, as $N \to \infty$ and $T \to \infty$ sequentially, is $O(T^{-1})$ and can be large.

Our approach centers on a calculation of Es_i, either analytically or numerically, and on the estimation of Es_i for given θ. Three mutually exclusive cases arise:

$$\begin{array}{rrcl} \text{(i)} & Es_i & = & 0; \\ \text{(ii)} & Es_i & \neq & 0 \text{ but } Es_i \text{ is free of } \eta_i; \\ \text{(iii)} & Es_i & \neq & 0 \text{ and } Es_i \text{ depends on } \eta_i. \end{array}$$

In case (i), s is unbiased and the MLE is consistent. The interesting point here is that a simple calculation, that of Es_i, will reveal so.

In case (ii), the MLE is inconsistent but the adjusted profile score $s - Es$ is unbiased and free of fixed effects. This paves the way for fixed-T consistent estimation. As it turns out, this is the case, surprisingly, in a number of static nonlinear models and in the linear dynamic model.

In case (iii), McCullagh and Tibshirani (1990) proposed using the adjusted profile score $s - \widehat{E}s$ instead of s, where $\widehat{E} = \widehat{E}_{\theta,\widehat{\eta}_i}$ is E but with $\widehat{\eta}_i$ replacing η_i.[1] The proposal was made in a more general context than the one considered here. McCullagh and Tibshirani discussed many examples, including several with incidental parameters, where the adjusted profile score improves on the profile score. In search of a general justification, they wrote (p. 342) "the centring of the profile log-likelihood function should improve the consistency of the maximizer of the likelihood" and yet, a few lines down, "We have no strong argument for this claim". We provide a large T asymptotic justification.

Let $T \to \infty$. Consider Es_i as a function of η_i. Under regularity conditions, replacing η_i with $\widehat{\eta}_i$ introduces a relative bias of $O(T^{-1})$, i.e. $E\widehat{E}s_i = (1 + O(T^{-1}))\,Es_i$ and, on averaging over i, $E\widehat{E}s = (1 + O(T^{-1}))\,Es$. Therefore, moving from the profile score s to McCullagh and Tibshirani's adjusted profile score $s - \widehat{E}s$ reduces the bias from $Es = O(T^{-1})$ to

$$E(s - \widehat{E}s) = O(T^{-2}).$$

That is, the adjustment removes the first-order bias from s, leaving only bias of order $O(T^{-2})$.

The adjustment can be iterated. The bias $E(s - \widehat{E}s)$ of the adjusted profile score can be appoximated by $\widehat{E}(s - \widehat{E}s)$, again with relative bias $O(T^{-1})$, and subtracted from $s - \widehat{E}s$ to give the second-order adjusted profile score

$$s - 2\widehat{E}s + \widehat{E}\widehat{E}s$$

[1] In addition to the centering step, McCullagh and Tibshirani also considered a rescaling of the adjusted profile score at to restore the information identity. We omit this step as our focus is on getting the estimating equation correctly centered. Similarly, an alternative adjustment would be to scale Es_i by the fisher information matrix, as in Firth (1993).

with bias

$$E(s - 2\widehat{E}s + \widehat{E}\widehat{E}s) = O(T^{-3}).$$

Note that, unlike $EEs = Es$ (since Es is a constant), $\widehat{E}\widehat{E}s \neq \widehat{E}s$ because $\widehat{E}\widehat{E}s$ is $E\widehat{E}s$ (a constant) but with E evaluated at $\widehat{\eta}_i$ instead of η_i for all i. The structure of the iterated adjustments is now apparent. Letting $\widehat{E}^{(k)}$ denote the $k$fold iteration of $\widehat{E}$, the jth order adjusted profile score is

$$\sum_{k=0}^{j} \binom{j}{k} (-1)^k \widehat{E}^{(k)} s,$$

with bias $O(T^{-j-1})$, given regularity conditions.

In general, the profile score adjustments, first and higher-order alike, have only an asymptotic justification. Note, however, that in case (ii) they all coincide with $s - Es$ and the adjustment is "exact". Whether there are interesting cases where some jth order adjustment is exact only as of some $j \geq 2$ is not known to us. In nonlinear models, the adjustments are generally not exact (though with important exceptions, as we shall see) but only approximations in the sense that they yield approximately unbiased estimating equations, to varying degrees of approximation. Nevertheless, it is hoped that, even when T is small, they yield improvements over the profile score. Whether that is true has to be examined on a case by case base. At the time of writing, our experience with the high-order adjustments is still limited, although we report on some simulations in Section 3.

Implementing the adjustments and solving the adjusted score equations requires evaluating Es_i for given θ and η_i. Often Es_i is not available in closed form, but it can be approximated by the average of R simulations of s_i. For large enough R, this average approximates Es_i to any desired accuracy, but for the sake of adjusting bias any R suffices, even $R = 1$. We do not recommend setting $R = 1$, however, except perhaps in models where evaluating s_i is computationally costly. A small R will only inflate the variance of the estimator somewhat. For the higher-order adjustments, which require evaluating terms like $\widehat{E}\widehat{E}s_i$, we suggest using small values of R in all inner expectations, and possibly a larger R in the outermost expectation. Finally, when approximating expectations by simulations, we suggest to keep the basic stream of random numbers used to generate R data sets, which is essentially of dimension $R \times N \times T \times \dim y_{it}$, constant for all values of θ and fixed effects, and for all levels of depth in $\widehat{E}^{(k)}$.

The adjustments discussed above seek to alter the estimating equation $s = 0$ to make it unbiased or approximately unbiased. Extensions are possible, perhaps even outside the parametric setting. One variation, in case (iii), is to slightly modify $s = 0$ so as make it, in essence, a case (ii) problem, where an exact adjustment is feasible. This possibility arises in the two-period logit model, as we discuss in Section 2. More generally, when $q = 0$ is some other estimating equation that is free of fixed effects and has bias $Eq = O(T^{-1})$, the type of adjustments discussed is possible provided that the expectation E can be evaluated. Another extension is to quantities other than θ, for example

$$\mu = \frac{1}{N}\sum_i \mu_i, \qquad \mu_i = \mu_i(\theta, \eta_i),$$

where $\mu_i(\cdot,\cdot)$ is a known function, such as a marginal effect. Replacing η_i with $\widehat{\eta}_i$ gives

$$m = \frac{1}{N}\sum_i m_i, \qquad m_i = m_i(\theta) = \mu_i(\theta, \widehat{\eta}_i),$$

with bias $O(T^{-1})$, given regularity conditions. The bias of m_i is $Em_i - \mu_i$ and can be appoximated as $\widehat{E}m_i - m_i$ to give $2m - \widehat{E}m$ as a first-order adjustment of m, and $\sum_{k=0}^{j}\binom{j+1}{k+1}(-1)^k \widehat{E}^{(k)} m$ as a jth order adjustment.

2 Examples

Our examples are models in which the distribution of a scalar y_{it} depends on a vector x_{it} through $\eta_i + \beta' x_{it}$ or $\eta_i \exp(\beta' x_{it})$. The common parameter consists of β and possibly an additional scale or shape parameter. Details about calculations of Es_i are given in the Appendix.

2.1 Models where $Es_i = 0$

There are several models with fixed effects but where there is no incidental-parameter problem.

Example 1 (Poisson counts) Consider Poisson counts y_{it} with mean $\lambda_{it} = \eta_i \exp(\beta' x_{it})$ and independence across t given $x_{i1}, ..., x_{iT}$. Here $f(y_{it}|x_{it}; \beta, \eta_i) = \exp(-\lambda_{it})\lambda_{it}^{y_{it}}/y_{it}!$ and $\theta = \beta$. Lancaster (2002) and Blundell, Griffith, and Windmeijer (2002) have shown that β and the η_i are likelihood orthogonal after a parameter transformation. Alternatively, a calculation shows that $Es_i = 0$. For given β, the MLE of η_i is $\widehat{\eta}_i = \sum_t y_{it} / \sum_t \exp(\beta' x_{it})$. Letting $\widehat{\lambda}_{it} = \widehat{\eta}_i \exp(\beta' x_{it})$, the profile log-likelihood and score for unit i are

$$\begin{aligned} l_i &= T^{-1}\sum_t(-\widehat{\lambda}_{it} + y_{it}\log\widehat{\lambda}_{it}) + \mathrm{c} \\ &= T^{-1}\sum_t y_{it}\left(-\log\sum_t \exp\left(\beta' x_{it}\right) + \beta' x_{it}\right) + \mathrm{c}, \\ s_i &= T^{-1}\sum_t y_{it}\left(-\frac{\sum_t \lambda_{it} x_{it}}{\sum_t \lambda_{it}} + x_{it}\right), \end{aligned}$$

where (here and later) c is an inessential constant. From $Ey_{it} = \lambda_{it}$, it follows that $Es_i = 0$. One may view $Es_i = 0$ as one implication of the conditional moment conditions given in Chamberlain (1993). The solution to $s = 0$, that is, the MLE, achieves the semiparametric efficiency bound (Hahn, 1997).

Example 2 (Exponential durations) Let y_{it} be exponentially distributed with mean λ_{it}^{-1}, where $\lambda_{it} = \eta_i \exp(\beta' x_{it})$, and independent across t given $x_{i1}, ..., x_{iT}$. The density is $f(y_{it}|x_{it};\beta,\eta_i) = \lambda_{it}\exp(-\lambda_{it}y_{it})$ and $\theta = \beta$. Here, $\widehat{\eta}_i = [T^{-1}\sum_t y_{it}\exp(\beta' x_{it})]^{-1}$. Again letting $\widehat{\lambda}_{it} = \widehat{\eta}_i \exp(\beta' x_{it})$, we have

$$
\begin{aligned}
l_i &= T^{-1}\sum_t (\log \widehat{\lambda}_{it} - \widehat{\lambda}_{it} y_{it}) \\
&= T^{-1}\sum_t \left(-\log \sum_t \left(y_{it}\exp\left(\beta' x_{it}\right)\right) + \beta' x_{it} \right) + \mathrm{c}, \\
s_i &= T^{-1}\sum_t \left(-\frac{\sum_t \lambda_{it} y_{it} x_{it}}{\sum_t \lambda_{it} y_{it}} + x_{it} \right).
\end{aligned}
$$

Conditionally on the x_{it}, the $\lambda_{it}y_{it}$ are i.i.d. unit-exponential variates. Therefore

$$
\begin{aligned}
E\frac{\sum_t \lambda_{it} y_{it} x_{it}}{\sum_t \lambda_{it} y_{it}} &= \sum_t \left(E\frac{\lambda_{it} y_{it}}{\sum_t \lambda_{it} y_{it}} \right) x_{it} = \sum_t \left(T^{-1} E\frac{\sum_t \lambda_{it} y_{it}}{\sum_t \lambda_{it} y_{it}} \right) x_{it} \\
&= T^{-1}\sum_t x_{it}
\end{aligned}
$$

and it follows that $Es_i = 0$. The MLE in the exponential regression model with fixed effects and exogenous regressors is fixed-T consistent. Greene's (2001) simulations support this, but a proof seems to be new.

2.2 Models where $Es_i \neq 0$ but Es_i is free of η_i

Example 3 (Many normal means) This is Neyman and Scott's (1948) classic example of the incidental-parameter problem. The problem is to infer $\theta = \sigma^2$ from independent observations $y_{it} \sim \mathcal{N}(\eta_i, \sigma^2)$. For any given σ^2, the MLE of η_i is $\overline{y}_i = T^{-1}\sum_t y_{it}$, so

$$
\begin{aligned}
l_i &= -(2T)^{-1}\sum_t (\log \sigma^2 + \sigma^{-2}(y_{it} - \overline{y}_i)^2), \\
s_i &= -(2T)^{-1}\sum_t (\sigma^{-2} - \sigma^{-4}(y_{it} - \overline{y}_i)^2),
\end{aligned}
$$

and $Es_i = -(2T\sigma^2)^{-1}$. The MLE of σ^2 is $(NT)^{-1}\sum_{i,t}(y_{it} - \overline{y}_i)^2$ and converges to $\sigma^2(1 - T^{-1})$. But since Es_i is free of η_i, a feasible unbiased estimating equation is $s - Es = 0$. Its root, $(NT - T)^{-1}\sum_{i,t}(y_{it} - \overline{y}_i)^2$, coincides with the outcome of many other approaches; see, e.g., McCullagh and Tibshirani (1990). A regression version of the model is $y_{it}|x_{i1}, ..., x_{iT} \sim \mathcal{N}(\eta_i + \beta x_{it}, \sigma^2)$, where θ now is β and σ^2. Here, the bias of the profile score for β is zero and for σ^2 it is $-(2T\sigma^2)^{-1}$, as before. Again, solving $s - Es = 0$ yields the standard solution: (i) the MLE of β (which is least-squares with unit-specific de-meaned data) is left unchanged; (ii) a one degree-of-freedom correction is applied to the MLE of σ^2. Even though the MLE of σ^2 is inconsistent and σ^2 and β are not profile likelihood orthogonal, the inconsistency does not carry over to the MLE of β. This is because the maximizer of $l(\beta, \sigma^2)$ with respect to β does not depend on σ^2 (though not vice versa).

Example 4 (Dynamic linear regression) Since Nickell (1981), dynamic linear models have become another classic instance of the incidental-parameter problem. Consider the model $y_{it} = \eta_i + \beta y_{it-1} + \varepsilon_{it}$ with $\varepsilon_{it} \sim \mathcal{N}(0, \sigma^2)$ and unrestricted initial observations y_{i0}. Let $y_i = (y_{i1}, ..., y_{iT})'$, $y_{i-} = (y_{i0}, ..., y_{iT-1})'$, and define the $T \times T$ matrix $M = I - T^{-1}\iota\iota'$ where ι is a vector of ones. For given β and σ^2, the MLE of η_i is $\widehat{\eta}_i = T^{-1}\iota'(y_i - \beta y_{i-})$. The profile log-likelihood and the elements of the profile score for unit i are

$$
\begin{aligned}
l_i &= -\frac{1}{2}\left(\log \sigma^2 + T^{-1}\sigma^{-2}(y_i - \beta y_{i-})' M (y_i - \beta y_{i-})\right) + \mathrm{c}, \\
s_{i\beta} &= -T^{-1}\sigma^{-2}(y_i - \beta y_{i-})' M y_{i-}, \\
s_{i\sigma^2} &= -\frac{1}{2}\left(\sigma^{-2} - T^{-1}\sigma^{-4}(y_i - \beta y_{i-})' M (y_i - \beta y_{i-})\right).
\end{aligned}
$$

Using backward substitution it is easy to show that

$$
E s_{i\beta} = -T^{-1}\sum_{t=1}^{T}(T-t)\beta^{t-1}, \qquad E s_{i\sigma^2} = -(2T\sigma^2)^{-1};
$$

see, e.g., Alvarez and Arellano (2004). Cox and Reid's (1987) orthogonalization approach leads to essentially the same result; see Lancaster (2002). While the adjusted score equation $s - Es = 0$ is unbiased it typically has more than one root, so the appropriate root has to be selected; see Dhaene and Jochmans (2010b). When the model is extended with exogenous covariates and p lags of y_{it}, Es_i is still available in closed form and remains free of η_i (Dhaene and Jochmans, 2010b). Finally, note that s_i depends only on the first two moments of the data, so the calculation of Es_i is robust to non-normality.

Example 5 (Weibull durations) In this model, y_{it}^{κ} is exponentially distributed with mean λ_{it}^{-1}, where $\lambda_{it} = \eta_i \exp(\beta' x_{it})$, and independent across t given $x_{i1}, ..., x_{iT}$. The density of y_{it} is $f(y_{it}|x_{it}; \beta, \kappa, \eta_i) = \kappa y_{it}^{\kappa-1}\lambda_{it}\exp(-\lambda_{it} y_{it}^{\kappa})$ and $\theta = (\beta', \kappa)'$. For given β and κ, the MLE of η_i is $\widehat{\eta}_i = [T^{-1}\sum_t y_{it}^{\kappa}\exp(\beta' x_{it})]^{-1}$. With $\widehat{\lambda}_{it} = \widehat{\eta}_i \exp(\beta' x_{it})$, the profile log-likelihood and score for unit i are

$$
\begin{aligned}
l_i &= T^{-1} \sum_t \left(\log \kappa + (\kappa - 1) \log y_{it} + \log \widehat{\lambda}_{it} - \widehat{\lambda}_{it} y_{it} \right) \\
&= T^{-1} \sum_t \left(\log \kappa + (\kappa - 1) \log y_{it} - \log \sum_t \left(y_{it}^{\kappa} \exp\left(\beta' x_{it} \right) \right) + \beta' x_{it} \right) + \text{c}, \\
s_{i\beta} &= T^{-1} \sum_t \left(- \frac{\sum_t y_{it}^{\kappa} \lambda_{it} x_{it}}{\sum_t y_{it}^{\kappa} \lambda_{it}} + x_{it} \right), \\
s_{i\kappa} &= T^{-1} \sum_t \left(\kappa^{-1} + \log y_{it} - \frac{\sum_t \left(\log y_{it} \right) y_{it}^{\kappa} \lambda_{it}}{\sum_t y_{it}^{\kappa} \lambda_{it}} \right).
\end{aligned}
$$

Given the x_{it}, the $\lambda_{it} y_{it}^{\kappa}$ are i.i.d., and so $Es_{i\beta} = 0$ by the same argument as in the exponential regression model. A calculation gives $Es_{i\kappa} = (\kappa T)^{-1}$, free of η_i, so the adjusted score $s - Es$ is feasible and unbiased. Although the profile score for β is unbiased, the MLE of β is inconsistent because the profile score for κ is biased and β and κ are not information orthogonal. Lancaster (2000) showed that an information-orthogonal transformation of η_i exists. Integrating the transformed effects from the likelihood using a uniform prior leads to Chamberlain's (1985) marginal-likelihood estimator.

Example 6 (Gamma durations) Here, y_{it} is gamma distributed with shape paramater κ and scale λ_{it}^{-1}, where $\lambda_{it} = \eta_i \exp\left(\beta' x_{it} \right)$, and independent across t given $x_{i1}, ..., x_{iT}$. The density function is $f(y_{it} | x_{it}; \beta, \kappa, \eta_i) = y_{it}^{\kappa - 1} \lambda_{it}^{\kappa} \exp\left(-\lambda_{it} y_{it} \right) / \Gamma(\kappa)$ and θ consists of β and κ. The MLE of η_i for given β and κ is $\widehat{\eta}_i = \kappa [T^{-1} \sum_t y_{it} \exp\left(\beta' x_{it} \right)]^{-1}$. Letting $\widehat{\lambda}_{it} = \widehat{\eta}_i \exp(\beta' x_{it})$ as before, the profile log-likelihood for unit i is

$$
l_i = T^{-1} \sum_t \left(- \log \Gamma(\kappa) + (\kappa - 1) \log y_{it} + \kappa \log \widehat{\lambda}_{it} - y_{it} \widehat{\lambda}_{it} \right)
$$

or, equivalently,

$$
T^{-1} \sum_t \left(- \log \Gamma(\kappa) + (\kappa - 1) \log y_{it} + \kappa \log\left(\kappa T \right) - \kappa - \kappa \log \sum_t \left(y_{it} \exp\left(\beta' x_{it} \right) \right) + \kappa \beta' x_{it} \right),
$$

with partial derivatives

$$
\begin{aligned}
s_{i\beta} &= T^{-1} \sum_t \kappa \left(- \frac{\sum_t y_{it} \lambda_{it} x_{it}}{\sum_t y_{it} \lambda_{it}} + x_{it} \right), \\
s_{i\kappa} &= T^{-1} \sum_t \left(-\psi(\kappa) + \log\left(\kappa T \right) + \log y_{it} - \log \sum_t \left(y_{it} \exp\left(\beta' x_{it} \right) \right) + \beta' x_{it} \right),
\end{aligned}
$$

where $\psi(\kappa)$ is the derivative of $\log \Gamma(\kappa)$. Again, $Es_{i\beta} = 0$, because the $y_{it} \lambda_{it}$ are independently gamma distributed with scale one and shape κ, given the x_{it}. A calculation shows that $Es_{i\kappa} =$

$\log(\kappa T) - \psi(\kappa T)$, again free of η_i. In this model, as in the linear model, there is an incidental-parameter problem only for κ. The solutions of $s = 0$ and $s - Es = 0$ differ for κ but coincide for β because, similar to the linear model, the maximizer of $l(\beta, \kappa)$ with respect to β does not depend on κ (though not vice versa). Using a similar argument as in the Weibull model, Chamberlain (1985) derived a fixed-T consistent estimator.

2.3 Models where $Es_i \neq 0$ and Es_i depends on η_i

Example 7 (Two-period negbin2 counts) In this model, y_{it} is a negbin2 count with mean $\lambda_{it} = \eta_i \exp(\beta' x_{it})$, variance $\lambda_{it} + \gamma^{-1}\lambda_{it}^2$, and there is independence across t given $x_{i1}, ..., x_{iT}$. The probability mass function for y_{it} is

$$f(y_{it}|x_{it}; \beta, \gamma, \eta_i) = \frac{\Gamma(\gamma + y_{it})}{\Gamma(\gamma)\Gamma(y_{it}+1)} \left(\frac{\lambda_{it}}{\lambda_{it}+\gamma}\right)^{y_{it}} \left(\frac{\gamma}{\lambda_{it}+\gamma}\right)^{\gamma},$$

where $\gamma > 0$ is an overdispersion parameter ($\gamma \to \infty$ yields Poisson counts). Unlike the fixed effects in the negbin1 model of Hausman, Hall, and Griliches (1984), the fixed effects enter the negbin2 model in the standard way, as a means to control for omitted time-invariant covariates; see, e.g., Allison and Waterman (2002) and Winkelmann (2008, p. 227–228). The common parameter is $\theta = (\beta', \gamma)'$. For $T = 2$, the simulation results in Allison and Waterman (2002) suggest that the MLE of β is free of incidental-parameter bias. The analysis below shows that there is incidental-parameter bias for β when $T = 2$, but that it is very small and can be ignored for practical purposes. For γ the incidental-parameter bias is much larger. For general T, the MLE of η_i for given β and γ satisfies

$$\sum_t \frac{y_{it} - \widehat{\eta}_i \exp(\beta' x_{it})}{\gamma + \widehat{\eta}_i \exp(\beta' x_{it})} = 0.$$

This equation is equivalent to a Tth order polynomial equation with a unique positive root. The uniqueness follows on rewriting the equation as

$$T^{-1} \sum_t \frac{y_{it} + \gamma}{\gamma + \widehat{\eta}_i \exp(\beta' x_{it})} = 1.$$

With $\widehat{\lambda}_{it} = \widehat{\eta}_i \exp(\beta' x_{it})$, the profile log-likelihood and score for unit i are

$$\begin{aligned}
l_i &= T^{-1} \sum_t \left(\log\Gamma(\gamma + y_{it}) - \log\Gamma(\gamma) + \gamma \log\gamma + y_{it} \log\widehat{\lambda}_{it} - (\gamma + y_{it}) \log(\gamma + \widehat{\lambda}_{it}) \right), \\
s_{i\beta} &= T^{-1} \sum_t \gamma \left(\frac{y_{it} - \widehat{\eta}_i \exp(\beta' x_{it})}{\gamma + \widehat{\eta}_i \exp(\beta' x_{it})} \right) x_{it}, \\
s_{i\gamma} &= T^{-1} \sum_t \left(\psi(\gamma + y_{it}) - \psi(\gamma) + \log\gamma - \log(\gamma + \widehat{\eta}_i \exp(\beta' x_{it})) \right).
\end{aligned}$$

The expectations of $s_{i\beta}$ and $s_{i\gamma}$ are

$$\begin{aligned} Es_{i\beta} &= \sum_{y_{i1}=0}^{\infty} \dots \sum_{y_{iT}=0}^{\infty} s_{i\beta} \prod_t f(y_{it}|x_{it};\beta,\gamma,\eta_i), \\ Es_{i\gamma} &= \sum_{y_{i1}=0}^{\infty} \dots \sum_{y_{iT}=0}^{\infty} s_{i\gamma} \prod_t f(y_{it}|x_{it};\beta,\gamma,\eta_i), \end{aligned}$$

but they are difficult to write in a more accessible form. We computed $Es_{i\beta}$ and $Es_{i\gamma}$ for $T=2$ and one-dimensional x_{it}. In this case, $\widehat{\eta}_i$ is the largest root of a quadratic equation and the expectations involve only double sums, so they can be evaluated fast and accurately. In general, $Es_{i\beta}$ and $Es_{i\gamma}$ are non-zero and depend on η_i. While $Es_{i\gamma}$ is large, $Es_{i\beta}$ is very small. We also computed $-(Eh_i)^{-1} Es_i$, where $h_i = \nabla_\theta s_i'$, as an approximation to the bias of the $(\widehat{\beta},\widehat{\gamma})'$, the MLE. When $Es_{i\beta}$ is small and, as turns out to be the case, $Eh_{i\beta\beta}$ is not too large and $Eh_{i\beta\gamma}$ is very small, this approximation to the bias of $\widehat{\beta}$ is accurate. We computed the approximate bias of $\widehat{\beta}$ over a range of values of γ and $x_{i1} < x_{i2}$, with η_i and β held fixed at one. For each γ, we found that the maximum (approximate) bias occurs as $x_{i1} \uparrow 0$ and $x_{i2} \downarrow 0$. Table 1 gives this maximum bias for γ corresponding to moderate to very high levels of overdispersion. Except for very small γ (very large overdispersion), the maximum bias is small. For other values of (b_1, b_2), the bias is typically much smaller than the maximum bias.

Table 1: Maximum approximate bias of $\widehat{\beta}$

γ	.5	1	2	5	10	20
maxbias$(\widehat{\beta})$	.1108	.0532	.0211	.0046	.0012	.0003

Table 2 shows Es and $E(s-\widehat{E}s)$ when $\eta_i = \beta = \gamma = 1$ and $x_{i1} = 0, x_{i1} = \log 2$, so that the means of y_{i1}, y_{i2} are $1, 2$ and the variances $2, 6$. For β the bias of the profile score is reduced by a factor 4, for γ by a factor 5.

Table 2: Bias before and after adjustment

	β	γ
Es	.0021	.133
$E(s-\widehat{E}s)$	.0005	.025

Example 8 (Two-period logit) Consider a pair $y_i = (y_{i1}, y_{i2})$ of independent variables y_{it} with mean $F(\eta_i + \beta x_{it})$, where $F(z) = (1+e^{-z})^{-1}$ is the logistic distribution at z and $(x_{i1}, x_{i2}) = (0,1)$ (see, e.g., Chamberlain, 1980). Here, $\theta = \beta$, the log-odds ratio, is the parameter

of interest. When y_i is $(0,0)$ or $(1,1)$ the MLE of η_i for any given β is infinite in absolute value and $l_i = s_i = 0$. For the movers, i.e., those units that have y_i equal to $(0,1)$ or $(1,0)$, the MLE is $\widehat{\eta}_i = -\beta/2$. Therefore, the profile loglikelihood and score for unit i are

$$\begin{aligned} l_i &= -d_{i01}\log(1+e^{-\beta/2}) - d_{i01}\log(1+e^{\beta/2}), \\ s_i &= \frac{1}{2}\left(\frac{d_{i01}}{1+e^{\beta/2}} - \frac{d_{i10}}{1+e^{-\beta/2}}\right), \end{aligned}$$

where d_{i01} is a binary indicator for $y_i = (0,1)$ and similarly for d_{i10}. Using

$$\begin{aligned} \pi_{i01} &= Ed_{i01} = \frac{1}{(1+e^{\eta_i})\,(1+e^{-\eta_i-\beta})}, \\ \pi_{i10} &= Ed_{i10} = \frac{1}{(1+e^{-\eta_i})\,(1+e^{\eta_i+\beta})} = \pi_{i01}e^{-\beta}, \end{aligned}$$

it follows that

$$Es_i = \frac{1}{2}\left(\frac{\pi_{i01}}{1+e^{\beta/2}} - \frac{\pi_{i01}e^{-\beta}}{1+e^{-\beta/2}}\right) = \frac{1}{2}\left(\frac{1-e^{-\beta/2}}{1+e^{\beta/2}}\right)\pi_{i01}.$$

Hence Es_i depends on η_i because π_{i01} does. Therefore, $s-\widehat{E}s$ is not unbiased (see also McCullagh and Tibshirani, 1990). However, a slight modification to s *before* addressing its bias essentially leads to case (ii), and to the conditional maximum-likelihood estimator (Andersen, 1970). Write s as

$$s = \frac{1}{2N}\left(\frac{N_{01}}{1+e^{\beta/2}} - \frac{N_{10}}{1+e^{-\beta/2}}\right),$$

where $N_{01} = \sum_i d_{i01}$ and N_{10} is defined similarly. Now consider

$$q = \frac{1}{2(N_{01}+N_{10})}\left(\frac{N_{01}}{1+e^{\beta/2}} - \frac{N_{10}}{1+e^{-\beta/2}}\right)$$

instead of s. Clearly, $s = 0$ and $q = 0$ have the same root—the MLE of β. If $\pi_{01} = \lim_{N\to\infty} N^{-1}\sum_i \pi_{i01}$ exists, then $\pi_{10} = \lim_{N\to\infty} N^{-1}\sum_i \pi_{i10} = \pi_{01}e^{-\beta}$ and

$$\begin{aligned} \text{plim}_{N\to\infty} q &= \frac{1}{2(\pi_{01}+\pi_{10})}\left(\frac{\pi_{01}}{1+e^{\beta/2}} - \frac{\pi_{10}}{1+e^{-\beta/2}}\right) \\ &= \frac{1-e^{-\beta/2}}{2(1+e^{-\beta})\,(1+e^{\beta/2})} = q_\infty \qquad \text{(say)}, \end{aligned}$$

where q_∞ is free of the sequence of η_i's. Therefore, $q - q_\infty = 0$ is a fixed-T unbiased estimating equation. Its solution coincides with the conditional maximum-likelihood estimator, which is known to be efficient (Hahn, 1997). While in this model there is nothing new in using $q - q_\infty = 0$, the illustration shows that normalizing the score function by the number of movers (or, more generally, by the number of informative units) can be helpful in models where conditioning is not possible.

3 Monte Carlo experiments

The upper panel in Figure 1 plots the profile score and various adjustments to it for the two-period logit model from Example 8. The plot was generated with $N = 1,000,000$, $\theta = 1$, and $\eta_i \sim \mathcal{N}(0, 1)$. The adjusted profile scores, shown up to third order, were obtained by means of one draw at each level of depth. The plot verifies the well-known results that, for this design, $\text{plim}_{N\to\infty}\widehat{\theta} = 2\theta$. The root of the first-order adjusted profile score is already much closer to the true parameter value. The higher-order corrections further reduce the inconsistency and the adjusted profile scores can be observed to converge to the score of the conditional-likelihood estimator as the order of correction increases. Also plotted is the (first-order) adjusted profile score using only units for which $y_{i1} + y_{i2} = 1$, that is, units that contribute to the profile likelihood. Its root is the true parameter value. As higher-order corrections would leave the location of this curve unchanged, these are not plotted.

The bottom panel in Figure 1 contains the scores for the probit variant of the two-period model. It may again be observed that a first-order adjustment greatly reduces the inconsistency of the MLE, and that iterating the corrections further centers the profile score. While, again, using only movers improves the situation, it does not lead to an adjusted profile score whose root is at the correct parameter value, although the inconsistency is small.

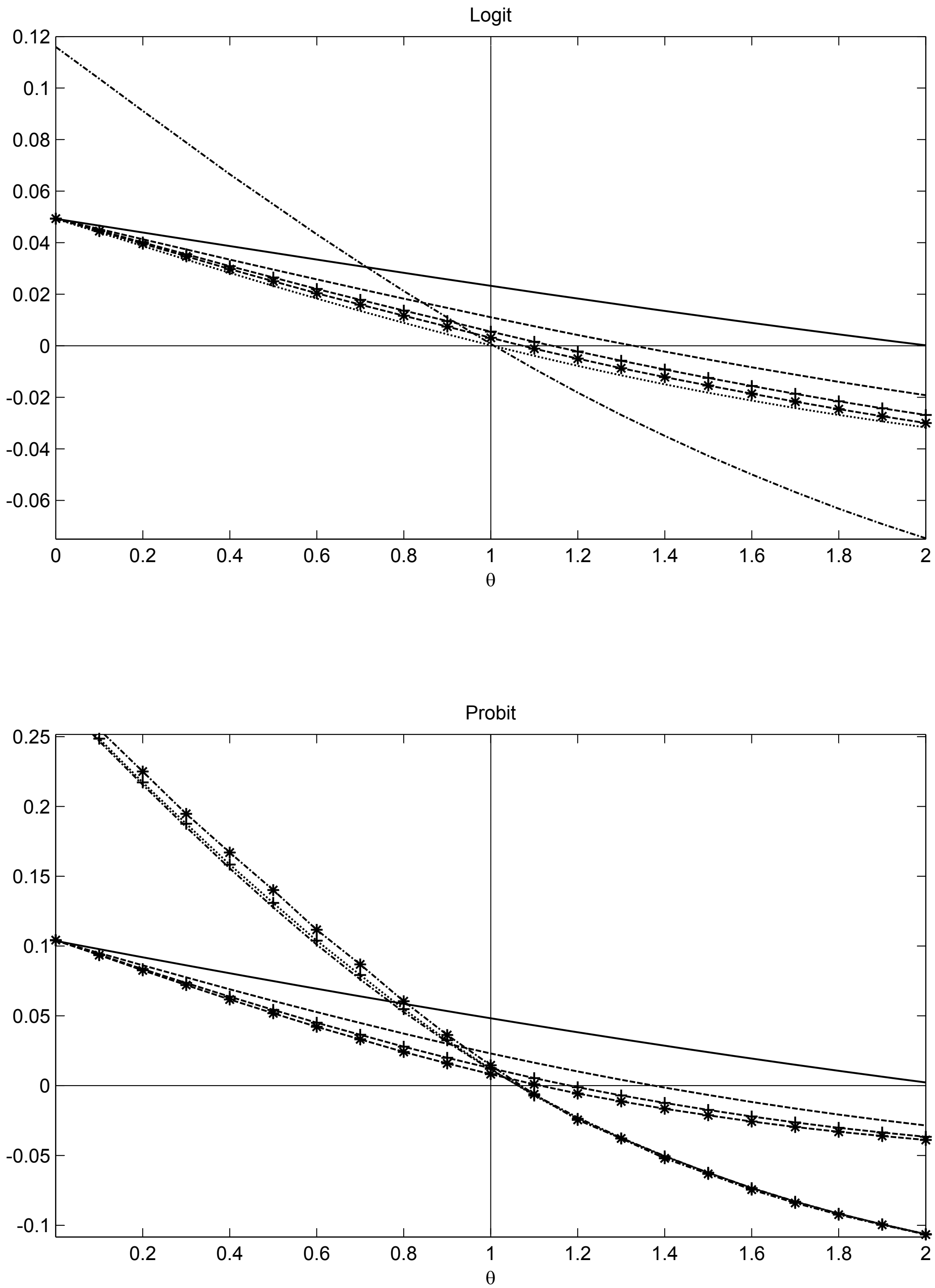
Figure 2 provides the profile score and the various adjustments for dynamic binary-choice models. They were obtained in an analogous fashion to before, only now with $T = 3$, which is the shortest panel length for which the MLE is finite. The start-up values for the N time series were drawn from their respective stationary distributions. The plots verify that dynamics tend to increase the magnitude of the inconsistency of the MLE. The re-centering effect of the score corrections is similar to before. The analogy of movers in the static model are units that alternate. More precisely, only units that switch status in each time period contribute to the profile likelihood. Here, using only such sequences does not lead to consistency of the root of the adjusted profile score in the logit model.

A small Monte Carlo experiment was performed to evaluate how much bias can be eliminated in small samples. The models considered are logit and probit variants of the binary-choice model

$$y_{it} = 1(\eta_i + x_{it}\theta \geq \varepsilon_{it}),$$

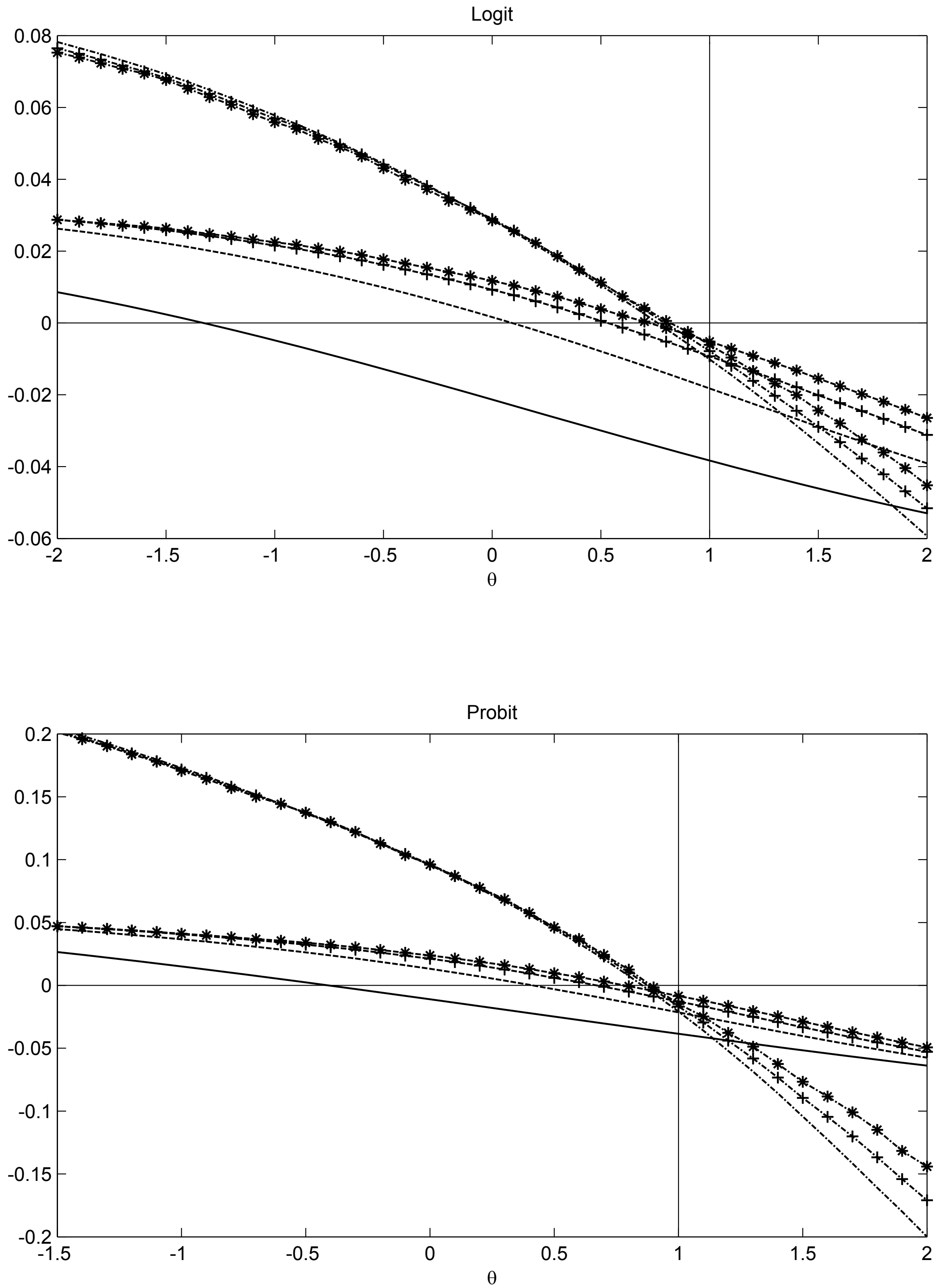
with both η_i and x_{it} scalar i.i.d. standard-normal variates. N was set to 100 and θ was fixed at unity throughout. Tables 3–4 contain the mean and median bias of the MLE $(\widehat{\theta})$ and the root of the first- and second-order corrected profile scores ($\widehat{\theta}_1$ and $\widehat{\theta}_2$), along with their standard deviation (STD) and interquartile range (IQR). The bias of the profile score was computed through simulation, with 10 runs for the outermost iteration step, and a single run in the inner iteration. For logit a useful benchmark is the conditional-likelihood estimator $(\widetilde{\theta})$, and so results for this estimator are also included. For probit, no fixed-T estimator is available.

Figure 1: Plots for static two-period binary-choice models



Score and adjusted score functions for the two-period logit (upper panel) and probit (lower panel) model with a time trend. $N = 1,000,000$; $\eta_i \sim \mathcal{N}(0,1)$. In both panels, $s(\theta)$ (solid) is plotted along with its first- (dashed), second- (dashed; marked $+$), and third-order (dashed; marked $*$) adjustments, together with their scaled counterparts (dashed–dotted). For the logit model, the score of the conditional likelihood (dotted) is also plotted.

Figure 2: Plots for dynamic three-period binary-choice models



Score and adjusted score functions for the three-period AR(1) logit (upper panel) and probit (lower panel) model. In both panels, $s(\theta)$ (solid) is plotted along with its first- (dashed), second- (dashed; marked $+$), and third-order (dashed; marked $*$) adjustments, together with their scaled counterparts (dashed–dotted).

Table 3: Results for the static logit model.

	MEAN BIAS				MEDIAN BIAS			
T	$\widehat{\theta}$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widetilde{\theta}$	$\widehat{\theta}$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widetilde{\theta}$
2	1.2080	.4275	.2294	.1040	1.0473	.3282	.1457	.0237
4	.4349	.0813	.0266	.0177	.4091	.0667	.0172	.0013
6	.2467	.0333	.0088	.0061	.2381	.0279	.0006	.0010
12	.1080	.0086	.0041	.0033	.1030	.0052	.0025	.0004
	STD				IQR			
T	$\widehat{\theta}$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widetilde{\theta}$	$\widehat{\theta}$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widetilde{\theta}$
2	.8198	.5231	.4879	.4099	.9205	.5723	.5393	.4603
4	.2687	.1870	.2005	.1754	.3532	.2435	.2638	.2353
6	.1740	.1380	.1556	.1331	.2370	.1884	.2144	.1821
12	.0964	.0869	.1019	.0853	.1324	.1226	.1406	.1189

Design: $N = 100$, $\theta = 1$, $\eta_i \sim \mathcal{N}(0,1)$, $x_{it} \sim \mathcal{N}(0,1)$, 250 replications.

Table 4: Results for the static probit model.

	MEAN BIAS			MEDIAN BIAS		
T	$\widehat{\theta}$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}$	$\widehat{\theta}_1$	$\widehat{\theta}_2$
2	1.0738	.3787	.1738	.9837	.3225	.1090
4	.5639	.1097	.0510	.5459	.0864	.0271
6	.3402	.0509	.0135	.3231	.04117	.0094
12	.1471	.0145	.0005	.1496	.0189	.0028
	STD			IQR		
T	$\widehat{\theta}$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}$	$\widehat{\theta}_1$	$\widehat{\theta}_2$
2	.5750	.3942	.3960	.6792	.4865	.4442
4	.2583	.1592	.1729	.3220	.2043	.2382
6	.1857	.1303	.1395	.2386	.1716	.1807
12	.0888	.0764	.0894	.1186	.0928	.1126

Design: $N = 100$, $\theta = 1$, $\eta_i \sim \mathcal{N}(0,1)$, $x_{it} \sim \mathcal{N}(0,1)$, 250 replications.

The tables show that the MLE suffers from a substantial upward bias. The results also suggest that our large-sample arguments to bias correction tend to give a reasonable approximation. Moreover, solving adjusted profile scores yields estimates with much smaller bias than the MLE. Iterating the correction further reduces the bias. It is apparent from the results that the bias is virtually fully eliminated when $T > 2$. Interestingly, the bias-corrected estimators are also less variable than is the MLE. It is known from Neyman and Scott (1948) that, with incidental parameters, the MLE need not be asymptotically efficient, even if it is fixed-T consistent. Notice, finally, that bias-corrected estimation of the logit model does not perform better than conditional-likelihood estimation.

Appendix

Weibull durations Write $\log y_{it}$ as

$$\log y_{it} = \kappa^{-1}\left(\log(y_{it}^{k}\lambda_{it}) - \log\lambda_{it}\right) = \kappa^{-1}\left(\log e_t - \log\lambda_{it}\right),$$

where the e_t are i.i.d. unit exponentials given the x_{it}. Then we have

$$s_{i\kappa} = T^{-1}\sum_t \left(\kappa^{-1} + \kappa^{-1}\left(\log e_t - \log\lambda_{it}\right) - \kappa^{-1}\frac{\sum_t\left(\log e_t - \log\lambda_{it}\right)e_t}{\sum_t e_t}\right),$$

with expectation

$$Es_{i\kappa} = \kappa^{-1}\left(1 + E\log e_t - TE\frac{e_t\log e_t}{\sum_t e_t}\right).$$

The sum in the denominator is $e_t + A$ where A is independent of e_t and has the Erlang distribution with density $A^{T-2}\exp(-A)/\,(T-2)!$. Therefore,

$$\begin{aligned} E\frac{e_t\log e_t}{\sum_t e_t} &= \int_0^\infty\int_0^\infty \frac{e\log e}{e+A}\exp(-e)\frac{A^{T-2}\exp(-A)}{(T-2)!}\;de\;dA \\ &= \frac{T-1-T\gamma}{T^2} \end{aligned}$$

where γ is Euler's gamma. We used Mathematica to calculate the integral. Setting $T = 1$ gives $E\log e_t = -\gamma$. On collecting results, $Es_{i\kappa} = (\kappa T)^{-1}$.

Gamma durations Here, write $\log y_{it}$ as

$$\log y_{it} = \log\left(y_{it}\lambda_{it}\right) - \log\eta_i - \beta' x_{it} = \log g_t - \log\eta_i - \beta' x_{it}$$

where, given the x_{it}, the g_t are i.i.d. gamma variates with shape κ and scale one. Then, write $s_{i\kappa}$ as

$$s_{i\kappa} = T^{-1}\sum_t\left(-\psi(\kappa) + \log\left(\kappa T\right) + \log g_t - \log\sum_t g_t\right).$$

Using $E\log g_t = \psi(\kappa)$ and the property that $\sum_t g_t$ is gamma distributed with shape κT and scale one, it follows that $Es_{i\kappa} = \log\left(\kappa T\right) - \psi(\kappa T)$.

References

Allison, P. D. and R. P. Waterman. (2002). Fixed-effects negative binomial regression models. *Sociological Methodology*, 32:247–265.

Alvarez, J. and M. Arellano. (2004). Robust likelihood estimation of dynamic panel data models. Unpublished manuscript.

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

Arellano, M. and S. Bonhomme. (2009) Robust priors in nonlinear panel data models. *Econometrica*, 77:489–536.

Arellano, M. and J. Hahn. (2006). A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects. Unpublished manuscript.

Arellano, M. and J. Hahn. (2007). Understanding bias in nonlinear panel models: Some recent developments. In R. W. Blundell, W. K. Newey, and P. Torsten (eds.), *Advances In Economics and Econometrics*, Volume III. Econometric Society, Cambridge University Press.

Arellano, M. and B. E. Honoré. (2001). Panel data models: Some recent developments. In J. J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, Volume VI. Elsevier.

Bester, C. A. and C. Hansen. (2009). A penalty function approach to bias reduction in non-linear panel models with fixed effects. *Journal of Business and Economic Statistics*, 27:131–148.

Blundell, R. W., R. Griffith, and F. Windmeijer. (2002). Individual effects and dynamics in count data models. *Journal of Econometrics*, 108:113–131.

Carro, J. M. (2007). Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics*, 140:503–528.

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

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

Chamberlain, G. (1993). Feedback in panel data models. Unpublished manuscript.

Chamberlain, G. (2010). Binary response models for panel data: Identification and information. *Econometrica*, 78:159–168.

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

Dhaene, G. and K. Jochmans. (2010a). Split-panel jackknife estimation of fixed-effect models. Unpublished manuscript.

Dhaene, G. and K. Jochmans. (2010b). An adjusted profile likelihood for non-stationary panel data models with fixed effects. Unpublished manuscript.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80:27–38.

Greene, W. H. (2001). Fixed and random effects in nonlinear models. Unpublished manuscript.

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

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

Hospido, L. (2010). Modelling heterogeneity and dynamics in the volatility of individual wages. *Journal of Applied Econometrics*, forthcoming.

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

Lancaster, T. (2002). Orthogonal parameters and panel data. *Review of Economic Studies*, 69:647–666.

Li, H., B. Lindsay, and R. Waterman. Efficiency of projected score methods in rectangular array asymptotics. *Journal of the Royal Statistical Society, Series B*, 65:191–208.

MacKinnon, J. G. and A. A. Smith Jr. (1998). Approximate bias correction in econometrics. *Journal of Econometrics*, 85:205–230.

McCullagh, P. and R. Tibshirani. (1990). A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society, Series B*, 52:325–344.

Mundlak, Y. (1961). Empirical production function free of management bias. *Journal of Farm Economics*, 43:44–56.

Neyman, J. and E. L. Scott. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16:1–32.

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

Sartori, N. (2003). Modified profile likelihood in models with stratum nuisance parameters. *Biometrika*, 90:533–549.

Winkelmann, R. (2008). *Econometric Analysis of Count Data*. Springer.