Integrated likelihoods in models with stratum nuisance parameters

Riccardo De Bin∗ Nicola Sartori† Thomas A. Severini‡

Abstract

Frequentist inference about a parameter of interest in presence of a nuisance parameter can be based on an integrated likelihood function. We analyze the behaviour of inferential quantities based on such a pseudo-likelihood in a two-index asymptotics framework, in which both sample size and dimension of the nuisance parameter may diverge to infinity. We show that a properly chose integrated likelihood largely outperforms standard likelihood methods, such as the profile likelihood. These results are confirmed by simulation studies, in which comparisons with modified profile likelihood are also considered.

Keywords: modified profile likelihood; non stationary autoregressive model; profile likelihood; profile score bias; target likelihood; two-index asymptotics.

∗debin@ibe.med.uni-muenchen.de - Department of Medical Informatics, Biometry and Epidemiology (IBE), Ludwig Maximilian University of Munich, Germany
†sartori@stat.unipd.it - Department of Statistical Sciences, University of Padova, Italy
‡severini@northwestern.edu - Department of Statistics, Northwestern University, USA
1 Introduction

Consider stratified data $y = (y_1, \ldots, y_q)$, where $y_i$ is a realization of an $m$-dimensional random variable $Y_i$ with density $p_i(\cdot; \psi, \lambda_i)$. Suppose that $Y_1, \ldots, Y_q$ are independent and consider $\psi$ as the parameter of interest, with $\lambda = (\lambda_1, \ldots, \lambda_q)$ as a nuisance parameter. We assume that each $\lambda_i$ has the same meaning and the same parameter space, $\Lambda$.

It is well known that, in models in which the dimension of the nuisance parameter is large relative to the sample size, methods of likelihood inference, such as those based on the profile likelihood, can perform poorly. To deal with this fact, several modifications to profile likelihood have been proposed; see Barndorff-Nielsen & Cox (1994, Chapter 8) and Severini (2000, Chapters 8, 9) for general discussion of these methods and further references. An alternative solution is offered by integrated likelihood functions (Kalbfleisch & Sprott, 1970), which are formed by integrating the likelihood function with respect to a weight function for the nuisance parameter. In our setting, a reasonable form for the weight function is $\prod_{i=1}^q g(\lambda_i; \psi)$, thus leading to an integrated likelihood

$$ L_I(\psi) = \prod_{i=1}^q \int_{\Lambda} p_i(y_i; \psi, \lambda_i) g(\lambda_i; \psi) \, d\lambda_i. $$

While this approach is quite common in a Bayesian framework, where $g(\lambda_i; \psi)$ is a conditional prior density of $\lambda$ given $\psi$, it is less used under the frequentist paradigm. A notable exception is inference on the index parameter of a composite group family, where a marginal likelihood for the index parameter can
be obtained from the distribution of the maximal invariant. Such marginal likelihood can be expressed in the form of an integrated likelihood (Pace & Salvan, 1997, Section 7.7.2). More generally, an integrated likelihood could be an attractive tool for inference since elimination of nuisance parameters is based on averaging, and this could avoid problems related to maximization that may arise with methods based on the profile likelihood, or its modifications (see, for example, Berger et al., 1999). From a frequentist point of view, inference on \( \psi \) proceeds by treating the integrated likelihood as a genuine likelihood for \( \psi \). In particular, we will study the properties of the usual quantities such as the integrated log likelihood \( l_I(\psi) = \log L_I(\psi) \), its maximizer \( \bar{\psi} \), and the likelihood ratio statistic, \( \bar{W} = 2\{l_I(\bar{\psi}) - l_I(\psi)\} \). When \( \psi \) is scalar, as often in the paper, also the integrated signed likelihood ratio statistic, \( \bar{R} = \text{sgn}(\bar{\psi} - \psi)\bar{W}^{\frac{1}{2}} \) will be considered.

In the following, we focus on frequentist properties of a special class of integrated likelihoods, as defined in Severini (2007, 2010, 2011), in models with many stratum nuisance parameters. Formally, we consider stratified models in which both \( m \), the within-stratum sample size, and \( q \), the number of strata, which is also the number of nuisance parameters, approach infinity. This type of “two-index asymptotics” is more relevant to cases in which the number of strata is large relative to the total sample size; see, e.g., Barndorff-Nielsen (1996) for a general discussion of two-index asymptotics and Sartori (2003) for discussion of the properties of profile and modified profile likelihoods in this setting.

A central role in the construction of the integrated likelihoods is played by the weight function \( g(\lambda_i; \psi) \). In a standard asymptotic setting, i.e. with
$q$ fixed, Severini (2007, 2010) shows the relative insensitivity of the integrated likelihood function to the choice of the weight function when the nuisance parameter is strongly unrelated to the parameter of interest and when such weight function does not depend on $\psi$. In this case, indeed, the integrated likelihood behaves asymptotically as a proper likelihood for $\psi$. Here we investigate the properties of the integrated likelihoods in the “two-index asymptotics” setting, providing both a specific class of weight functions and a condition on $m$ and $q$ that guarantee the usual asymptotic behaviour of likelihood quantities based on the integrated likelihood. For instance, we show that, using one of these weight functions, $\bar{R}$ has an asymptotic standard normal distribution even in extreme settings, i.e. large $q$ and small $m$, when the same results do not hold for $R$, the usual signed likelihood ratio statistic based on the profile likelihood. The results for the integrated likelihood are equivalent to those for the modified profile likelihood in Sartori (2003). We give however some examples in which the integrated likelihood outperforms the modified profile likelihood, thus suggesting that it may be a useful alternative.

Although the present paper does not deal with Bayesian inference, we note that the proposed integrated likelihood may be used in conjunction with a prior for $\psi$ only in order to obtain an approximate marginal posterior distribution for $\psi$, as suggested for instance in Efron (1993) and Ventura et al. (2009). In particular, Ventura et al. (2009) propose a valid matching prior for $\psi$ that can be used with the modified profile likelihood for $\psi$. Given the connection between the integrated likelihood used here and the modified profile likelihood, the same matching prior could be used also with the inte-
grated likelihood in order to obtain Bayesian inference with good frequentist properties.

The paper is organized as follows: in Section 2 we describe the target likelihood, an unavailable ideal likelihood for \( \psi \), that we approximate using a suitable integrated likelihood; the latter is based on an appropriate choice of the weight function, which is discussed in Section 3. The asymptotic properties of the likelihood quantities based on the integrated likelihoods, such as the integrated score function, the maximum integrated likelihood estimator and the integrated likelihood ratio statistics, are studied in Section 4. Examples are presented in Section 5, in which comparisons with profile and modified profile likelihoods are considered.

## 2 Target likelihood

Let \( L(\psi, \lambda) = \prod_{i=1}^{q} p_i(y_i; \psi, \lambda_i) \) denote the likelihood function and \( l(\psi, \lambda) = \sum_{i=1}^{q} \log p_i(y_i; \psi, \lambda_i) = \sum_{i=1}^{q} l_i(\psi, \lambda_i; y_i) \) denote the log-likelihood function. Let \( \lambda_{i\psi} \) denote the maximum likelihood estimator of \( \lambda_i \) for fixed \( \psi \) and let \( l_P(\psi) = \sum_{i=1}^{q} l_i(\psi, \hat{\lambda}_{i\psi}; y_i) \) denote the profile log-likelihood. Derivatives of the log-likelihood will be denoted by subscripts; e.g., \( l_{\lambda_i}(\psi, \lambda_i) = \partial l(\psi, \lambda) / \partial \lambda_i = \partial l_i(\psi, \lambda_i) / \partial \lambda_i \); \( j_{\psi\psi}(\psi, \lambda) \) and \( j_{\lambda\lambda}(\psi, \lambda) \) will denote blocks of the observed information matrix. Similarly, \( i_{\psi\psi}(\psi, \lambda) \) and \( i_{\lambda\lambda}(\psi, \lambda) \) will denote blocks of the expected information matrix. Note that, due to independence among the strata, \( j_{\lambda\lambda}(\psi, \lambda) \) and \( i_{\lambda\lambda}(\psi, \lambda) \) are block diagonal matrices, with generic element \( j_{\lambda_i\lambda_i}(\psi, \lambda_i) \) and \( i_{\lambda_i\lambda_i}(\psi, \lambda_i) \) respectively. For notational simplicity, we will consider scalar nuisance parameters; the results are easily extended.
to the case in which the $\lambda_i$ are vectors.

We will study the asymptotic properties of integrated likelihoods by relating them to the least favourable target likelihood (Pace & Salvan, 2006). Let $E_0$ denote expectation with respect to the true parameter $(\psi_0, \lambda_0)$, where $\lambda_0 = \lambda_{10}, \ldots, \lambda_{q0}$. Define $\lambda_{i\psi} \equiv \lambda_i(\psi; \psi_0, \lambda_{i0})$ as the maximizer of $E_0[l_i(\psi, \lambda_i)]$ in $\lambda_i$ for fixed $\psi$; the target log-likelihood is given by $l_T(\psi) = \sum_{i=1}^{q} l_i(\psi, \lambda_{i\psi})$.

The target likelihood is a function of $\psi$, the data, and the true parameter value $(\psi_0, \lambda_0)$; it is a genuine likelihood for $\psi$, but it is not available in practice since it depends on $(\psi_0, \lambda_0)$. In some sense, it is analogous to a “true value” for a likelihood for a parameter of interest. Note that likelihood quantities based on the target likelihood have the usual asymptotic properties of functions of a likelihood for the parameter $\psi$; for instance, the target score, $l_{T\psi}$, divided by $\sqrt{mq}$, has an asymptotic normal distribution with mean 0 and variance equal to the expected information, regardless of the rates of divergence of $q$ and $m$. Thus, relating statistical procedures based on an integrated likelihood to the ones based on the target likelihood is a useful and convenient way to establish asymptotic results in complex settings such as the one considered here. The target likelihood is also closely related to the modified profile likelihood and related pseudolikelihood functions (Pace & Salvan, 2006).

Exact integration of the likelihood function is possible only in exceptional cases; in all other cases integration in practice is performed either numerically or relying on analytical approximations. In deriving the properties of procedures based on the integrated likelihood, here we follow the latter approach, exploiting the features of the Laplace approximations. Since each $\lambda_i$ appears
only in a single stratum, the analytical approximation of \( l_I(\psi) = \log L_I(\psi) \) can be obtained by using a Laplace approximation in each stratum and then combining the results,

\[
l_I(\psi) = l_P(\psi) + \sum_{i=1}^{q} \log g(\hat{\lambda}_i\psi; \psi) - \sum_{i=1}^{q} \log j_{\lambda_i\lambda_i}(\psi, \hat{\lambda}_i\psi)^{1/2} + O_p(q/m). \tag{1}
\]

Expansion of the target likelihood (Pace & Salvan, 2006) yields

\[
l_T(\psi) = l_P(\psi) + \frac{1}{2} \sum_{i=1}^{q} (\hat{\lambda}_i\psi - \lambda_i\psi)^2 l_{\lambda_i\lambda_i}(\psi, \lambda_i\psi) - \frac{1}{6} \sum_{i=1}^{q} (\hat{\lambda}_i\psi - \lambda_i\psi)^3 l_{\lambda_i\lambda_i\lambda_i}(\psi, \lambda_i\psi) + O_p(q/m).
\]

The second summand of the right hand side of this formula can be approximated (Pace & Salvan, 2006) by

\[
-\frac{1}{2} \sum_{i=1}^{q} \log(\hat{\lambda}_i\psi - \lambda_i\psi)^2 - \frac{1}{2} \sum_{i=1}^{q} \log j_{\lambda_i\lambda_i}(\psi, \lambda_i\psi),
\]

while the third summand is a term of order \( O_p(\max\{\sqrt{q/m}, q/m\}) \). The latter, rather unconventional, expression is required by the unconventional two-index asymptotics, in which \( q \) and \( m \) may diverge at different rates (Sartori, 2003). Hence

\[
l_T(\psi) = l_P(\psi) - \frac{1}{2} \sum_{i=1}^{q} \log(\hat{\lambda}_i\psi - \lambda_i\psi)^2 - \frac{1}{2} \log j_{\lambda_i\lambda_i}(\psi, \lambda_i\psi) + O_p(\max\{\sqrt{q/m}, q/m\}). \tag{2}
\]

Combining (1) and (2),

\[
l_I(\psi) = l_T(\psi) + \sum_{i=1}^{q} \log g(\hat{\lambda}_i\psi; \psi) + \frac{1}{2} \sum_{i=1}^{q} \log(\hat{\lambda}_i\psi - \lambda_i\psi)^2 + O_p(\max\{\sqrt{q/m}, q/m\}). \tag{3}
\]
This equation can be used to relate quantities based on the integrated likelihood to the corresponding quantities based on the target likelihood.

### 3 Target weight functions

According to (3), \( l_I(\psi) = l_T(\psi) + O(q) \) in general. To reduce the order of the error term in this relationship, we can choose a weight function \( g \) such that

\[
\log g(\hat{\lambda}_i; \psi) + \frac{1}{2} \log(\hat{\lambda}_i - \lambda_i)^2 = O_p(1/m);
\]

under this condition, \( l_I(\psi) = l_T(\psi) + O_p(\max\{\sqrt{q/m}, q/m\}) \). We will refer to weight functions which satisfy this requirement as **target weight functions**.

From the expansion of the term \((\hat{\lambda}_i - \lambda_i)\), using the results of Pace & Salvan (2006), it is possible to show that functions of the form

\[
g(\lambda; \psi) = \frac{j_{\lambda,\lambda}(\psi, \lambda)}{\nu_{\lambda,\lambda}((\psi, \lambda), (\psi, \lambda); (\hat{\psi}, \hat{\lambda}))^{1/2}}
\]
or

\[
g(\lambda; \psi) = \frac{j_{\lambda,\lambda}(\psi, \lambda)}{\nu_{\lambda,\lambda}((\psi, \lambda), (\psi, \lambda); (\hat{\psi}, \hat{\lambda}))},
\]

fulfill (4) and therefore are example of target weight functions. Here \( \nu_{\theta_0, \theta_1}(\theta_0, \theta_1; \theta_2) = E_{\theta_2}[l_{\theta}(\theta_0)l_{\theta}(\theta_1)] \).

The derivation of these functions, however, may be cumbersome. It may be easier, therefore, to rely on an orthogonal parameterization to construct an integrated likelihood satisfying \( l_I(\psi) = l_T(\psi) + O_p(\max\{\sqrt{q/m}, q/m\}) \).

If \( \lambda_i \) and \( \psi \) are orthogonal parameters, indeed, then \( \hat{\lambda}_i = \hat{\lambda}_i + O(1/m) \) and
If \( \lambda_i \) is orthogonal to \( \psi \), therefore, any weight function for \( \lambda_i \) not depending on \( \psi \) is a target weight function, and the integrated likelihood agrees to the target likelihood up to order \( O_p(\max\{\sqrt{q/m}, q/m\}) \). In particular, the weight function \( g(\lambda; \psi) = 1 \) is useful in this context; that is, it is sufficient to integrate the likelihood with respect to \( \lambda \) to obtain the required integrated likelihood.

Here, the orthogonal parameter can be taken to be the information-orthogonal parameter discussed in detail by Cox & Reid (1987), and obtained by solving a differential equation based on the expected information matrix. Alternatively, it may be taken to be the zero-score expectation (ZSE) parameter used in Severini (2007), which is obtained by solving

\[
E[\ell_{\lambda_i}(\psi, \lambda_i); \psi_0, \phi_i] = 0
\]  

(5)

to obtain expression for \( \phi_i \) in terms of \( \psi, \lambda_i, \psi_0 \); \( \psi_0 \) is then replaced by an estimator, such as the maximum likelihood estimator.

The integrated likelihood based on a target weight function for the ZSE parameter has the advantage that it approximately satisfies the second Bartlett identity; this is not true if the integrated likelihood is based on a target weight function for an information-orthogonal parameter (Severini, 2007). One consequence of this is that, in some cases, inferences based on an integrated likelihood using a target weight function for the ZSE parameter are preferable to inferences based on an integrated likelihood using a target weight function
for the information-orthogonal parameter. On the other hand, the ZSE pa-
rameter requires a reliable estimator of $\psi$, which may not be available in the
setting considered here. Also, for some models the information-orthogonal
parameter is easier to obtain, while in other models the reverse is true. Thus,
both approaches are useful in practice; this is illustrated in the examples.

4 Asymptotic properties of some integrated
likelihood-based quantities

4.1 Score function

First consider the relationship between the score function based on an inte-
grated likelihood function and the score function based on the target likeli-
hood. Using (3),

$$l_{I\psi}(\psi) = l_{T\psi}(\psi) + \sum_{i=1}^{q} \frac{\partial}{\partial \psi} \log g(\hat{\lambda}_{i\psi}; \psi) + \frac{1}{2} \sum_{i=1}^{q} \frac{\partial}{\partial \psi} \log(\hat{\lambda}_{i\psi} - \lambda_{i\psi})^2$$

+ $O_p(\max\{\sqrt{q/m}, q/m\})$.

The term

$$D(\psi) = \sum_{i=1}^{q} \frac{\partial}{\partial \psi} \log g(\hat{\lambda}_{i\psi}; \psi) + (1/2) \sum_{i=1}^{q} \frac{\partial}{\partial \psi} \log(\hat{\lambda}_{i\psi} - \lambda_{i\psi})^2$$

is in general of order $O_p(q)$. If $g$ is a target weight function, then $D(\psi) =
O_p(q/m)$ and the discrepancy between the two score functions is of order
$O_p(\max\{q/m, \sqrt{q/m}\})$.  

10
Since, by definition, the target score is unbiased, this result allows us to study the bias of the integrated score. In general, \( E[D(\psi)] = O(q) \), which means that \( l_I(\psi) \) has score bias of order \( O(q) \); this is the same order as the score bias of the profile likelihood (Sartori, 2003). With a target weight function, the order of the integrated score bias is \( O(q/m) \), that is the same as the order of the score bias of the modified profile likelihood (Sartori, 2003).

4.2 Maximum integrated likelihood estimator

Let \( \tilde{\psi} \) and \( \hat{\psi}_T \) denote the maximizers of \( l_I(\psi) \) and \( l_T(\psi) \), respectively. Note that \( \sqrt{nq}(\hat{\psi}_T - \psi) \) is asymptotically normally distributed, with 0 mean and variance equal to the inverse of the partial expected information for \( \psi \)

\[
i_{\psi;\lambda}(\psi, \lambda) = i_{\psi\psi}(\psi, \lambda) - i_{\psi\lambda}(\psi, \lambda)i_{\lambda\lambda}(\psi, \lambda)^{-1}i_{\lambda\psi}(\psi, \lambda),
\]

which is the lower bound for the asymptotic covariance matrix of a regular estimator of \( \psi \) when \( \lambda \) is unknown (Bahadur, 1964).

Recall that the maximizer \( \hat{\psi} \) of a log-likelihood \( l(\psi) \) based on \( n \) observations can be expanded (e.g., Severini, 2000, Section 5.3)

\[
\sqrt{n}(\hat{\psi} - \psi) = -\frac{Z_1}{\mu_{\psi\psi}} + \left( \frac{Z_1Z_2}{\mu_{\psi\psi}} - \frac{1}{2} \frac{\mu_{\psi\psi}}{\mu_{\psi\psi}} Z_1 \right) \frac{1}{\sqrt{n}} + O_p \left( \frac{1}{n} \right),
\]

where

\[
Z_1 = \frac{l_\psi(\psi) - E_\psi[l_\psi(\psi)]}{\sqrt{n}}, \quad Z_2 = \frac{l_{\psi\psi}(\psi) - E_\psi[l_{\psi\psi}(\psi)]}{\sqrt{n}},
\]

\[
\mu_{\psi\psi} = \frac{1}{n} E_\psi[l_{\psi\psi}(\psi)], \quad \mu_{\psi\psi\psi} = \frac{1}{n} E_\psi[l_{\psi\psi\psi}(\psi)].
\]
Applying this expansion to $\bar{\psi}$ and $\hat{\psi}_T$, we have

$$\sqrt{mq}(\bar{\psi} - \psi) = -\frac{\bar{Z}_1}{\bar{\mu}_{\psi\psi}} + \left(\frac{\bar{Z}_1 \bar{Z}_2}{\bar{\mu}_{\psi\psi}^2} - \frac{1}{2} \frac{\bar{\mu}_{\psi\psi}}{\bar{\mu}_{\psi\psi}^3} \bar{Z}_1^2\right) \frac{1}{\sqrt{mq}} + O_p\left(\frac{1}{mq}\right),$$

$$\sqrt{mq}(\hat{\psi}_T - \psi) = -\frac{\tilde{Z}_1}{\tilde{\mu}_{\psi\psi}} + \left(\frac{\tilde{Z}_1 \tilde{Z}_2}{\tilde{\mu}_{\psi\psi}^2} - \frac{1}{2} \frac{\tilde{\mu}_{\psi\psi}}{\tilde{\mu}_{\psi\psi}^3} \tilde{Z}_1^2\right) \frac{1}{\sqrt{mq}} + O_p\left(\frac{1}{mq}\right),$$

where the symbol $\bar{\cdot}$ denotes quantities based on $l_I(\psi)$, while $\tilde{\cdot}$ denotes quantities based on $l_T(\psi)$.

Since, in general, the following relations hold

$$l_{I\psi}(\psi) = l_{T\psi}(\psi) + O_p(q), \quad l_{I\psi\psi}(\psi) = l_{T\psi\psi}(\psi) + O_p(q),$$

$$\bar{Z}_1 = \tilde{Z}_1 + O_p(\sqrt{q/m}), \quad \bar{Z}_2 = \tilde{Z}_2 + O_p(\sqrt{q/m}),$$

$$\bar{\mu}_{\psi\psi} = \tilde{\mu}_{\psi\psi} + O_p(1/m), \quad \bar{\mu}_{\psi\psi\psi} = \tilde{\mu}_{\psi\psi\psi} + O_p(1/m),$$

then we have

$$\frac{\bar{Z}_1}{\bar{\mu}_{\psi\psi}} = \frac{\tilde{Z}_1}{\tilde{\mu}_{\psi\psi}} + O_p(\sqrt{q/m}) + O_p(1/m)$$

and

$$\frac{\bar{Z}_1 \bar{Z}_2}{\bar{\mu}_{\psi\psi}^2} - \frac{1}{2} \frac{\bar{\mu}_{\psi\psi}}{\bar{\mu}_{\psi\psi}^3} \bar{Z}_1^2 = \frac{\tilde{Z}_1 \tilde{Z}_2}{\tilde{\mu}_{\psi\psi}^2} - \frac{1}{2} \frac{\tilde{\mu}_{\psi\psi}}{\tilde{\mu}_{\psi\psi}^3} \tilde{Z}_1^2 + O_p(1/m) + O_p(\max\{\sqrt{q/m}, q/m\}).$$

Hence, in general, we have

$$\sqrt{mq}(\bar{\psi} - \psi) = \sqrt{mq}(\hat{\psi}_T - \psi) + O_p(1/m) + O_p(\max\{\sqrt{q/m}, q/m\}).$$
On the other hand, for an integrated likelihood based on a target weight function, the analogous relations are

\[
l_I'(\psi) = l_{T\psi}(\psi) + O_p(q/m), \quad l_{T\psi}(\psi) = l_{T\psi}(\psi) + O_p(q),
\]

\[
\bar{Z}_1 = \tilde{Z}_1 + O_p(\sqrt{q/m^3}), \quad \bar{Z}_2 = \tilde{Z}_2 + O_p(\sqrt{q/m})
\]

\[
\hat{\mu}_{\psi\psi} = \hat{\mu}_{\psi\psi} + O_p(1/m), \quad \hat{\mu}_{\psi\psi\psi} = \hat{\mu}_{\psi\psi\psi} + O_p(1/m),
\]

which lead to

\[
\frac{\bar{Z}_1}{\hat{\mu}_{\psi\psi}} = \frac{\tilde{Z}_1}{\mu_{\psi\psi}} + O_p(\sqrt{q/m^3}) + O_p(1/m).
\]

and

\[
\frac{\bar{Z}_1 \bar{Z}_2 - \frac{1}{2} \frac{\hat{\mu}_{\psi\psi}}{\mu_{\psi\psi}} \bar{Z}_1^2}{\hat{\mu}_{\psi\psi}} = \frac{\tilde{Z}_1 \tilde{Z}_2 - \frac{1}{2} \frac{\hat{\mu}_{\psi\psi}}{\mu_{\psi\psi}} \tilde{Z}_1^2}{
\mu_{\psi\psi}^2} + O_p(\sqrt{q/m}) + O_p(1/m) + O_p(\max\{\sqrt{q/m^3}, q/m^3\})
\]

Hence,

\[
\sqrt{mq} (\tilde{\psi} - \psi) = \sqrt{mq} (\hat{\psi}_T - \psi) + O_p(1/m) + O_p(\sqrt{q/m^3}).
\]

It follows that, in general, \(\tilde{\psi}\) has the same asymptotic distribution as \(\hat{\psi}_T\) and, hence, is asymptotically efficient, provided that \(q/m = o(1)\). When the integrated likelihood is based on a target weight function, instead, \(\tilde{\psi}\) has the same asymptotic distribution as \(\hat{\psi}_T\) and it is asymptotically efficient provided that \(q/m^3 = o(1)\).
4.3 Integrated likelihood ratio statistics

We now consider likelihood-ratio-type statistics based on the integrated likelihood. Since the target likelihood is a likelihood for $\psi$ based on a sample of size $mq$, the likelihood ratio statistic for $\psi$ based on the target likelihood, $W_T$, is asymptotically distributed according to a chi-squared distribution with $p$ degrees-of-freedom, where $p = \dim(\bar{\psi})$. Using the relationships between derivatives of $l_I$ and derivatives of $l_T$, along with the relationship between $\bar{\psi}$ and $\hat{\psi}_T$, it is straightforward to show that

$$l_I(\bar{\psi}) - l_I(\psi) = l_T(\hat{\psi}_T) - l_T(\psi) + \Delta_{q,m}$$

where, in general, $\Delta_{q,m} = O_p(\sqrt{q/m})$ and for a target weight function, $\Delta_{q,m} = O_p(1/\sqrt{m}) + O_p(\sqrt{q/m^3})$.

Let $\bar{W}$ denote the likelihood ratio statistic for $\psi$ based on the integrated likelihood. It follows that, in general,

$$\bar{W} = W_T + \Delta_{q,m}$$

and, hence, that $\bar{W}$ is in general asymptotically chi-squared-distributed, provided that $q/m = o(1)$. For an integrated likelihood based on a target weight function, this condition is weakened to $q/m^3 = o(1)$.

The same approach can be used for signed likelihood ratio statistics for a scalar parameter $\psi$. Using the results above, it is easily shown that the signed integrated likelihood ratio statistic $\bar{R}$ is approximately equal to $R_T$, the signed likelihood ratio statistic based on the target likelihood, with error
$\Delta_{q,m}$. Since $R_T$ is asymptotically standard normal with error $1/\sqrt{mq}$ it follows that, in general, $\bar{R}$ is asymptotically standard normal provided that $q/m = o(1)$; again, if the integrated likelihood is based on a target weight function, the required condition is $q/m^3 = o(1)$.

The target weight function used to form the integrated likelihood can be based on either the ZSE parameter or on an information-orthogonal parameter and the properties of $\bar{R}$ discussed in this section hold in either case. However, there is a sense in which the properties of $\bar{R}$ are better if a ZSE-based integrated likelihood is used. Specifically, in this case, the standard deviation of $\bar{R}$ can be expanded as $1 + O(1/m^2) + O(1/(mq))$ while for the information-orthogonal-based integrated likelihood, the standard deviation of $\bar{R}$ can be expanded as $1 + O(1/m) + O(1/(mq))$. However, in both cases, $E[\bar{R}]$ can be expanded as $O(\sqrt{q/m^3}) + O(1/\sqrt{mq})$ and these terms will often be more important than the error in the standard deviation. Moreover, issues related to the ease in solving for the parameter and the necessity of finding a useful estimator of $\psi$ to use in forming the ZSE parameter, as discussed in Section 2, typically play a more important role in choosing a parameterization.

The results above for the integrated likelihoods can be compared to those for likelihood ratio statistics based on the profile and modified profile likelihoods. For example, Sartori (2003) shows that, if $q/m = o(1)$, then the usual signed likelihood ratio statistic is asymptotically normally distributed and that the signed likelihood ratio statistic based on the modified profile likelihood is asymptotically normal provided that $q/m^3 = o(1)$. Hence, from a theoretical point of view, an integrated likelihood based on a target weight
function guarantees the same inferential accuracy as the modified profile likelihood. On the other hand, from a practical point of view, there may be instances in which the integrated likelihood approach may be preferable, both in terms of accuracy and computational requirements, as illustrated in some of the examples below. Another example is small sample meta-analysis, as described in Bellio & Guolo (2014).

5 Examples

Example 1: gamma samples with common shape parameter. Let $Y_{ij}$, $i = 1, \ldots, q$, $j = 1, \ldots, m$, be independent gamma random variables with shape parameter $\psi$ and scale parameter $1/\lambda_i$, as in Sartori (2003, Example 2). Writing $s = u - m \sum_{i=1}^{q} \log v_i$, where $u = \sum_{i=1}^{q} \sum_{j=1}^{m} \log y_{ij}$ and $v_i = \sum_{j=1}^{m} y_{ij}$ are the components of the sufficient statistic, the conditional and profile log-likelihoods are

$$l_C(\psi) = \psi s + q \log \Gamma(m\psi) - mq \log \Gamma(\psi),$$

$$l_P(\psi) = \psi s + mq \psi \log m\psi - mq\psi - mq \log \Gamma(\psi),$$

while, if we use as a weight function the density function of an exponential random variable with mean 1, the integrated log-likelihood is

$$l_I(\psi) = \psi \sum_{i=1}^{q} \sum_{j=1}^{m} \log y_{ij} - m\psi \sum_{i=1}^{q} \log(\sum_{j=1}^{m} y_{ij} + 1) + q \log \Gamma(m\psi + 1)$$

$$-mq \{\psi + \log \Gamma(\psi)\}. \quad (6)$$
Table 1: Example 1. Empirical coverage for $R$, $R_C = \bar{R}_I$, $\bar{R}_G$ and $R_{MP}$.

| nominal | $R$  | $R_C = \bar{R}_I$ | $\bar{R}_G$ | $R_{MP}$ |
|---------|------|-------------------|-------------|----------|
| 0.01    | 0.000| 0.010             | 1.000       | 0.010    |
| 0.025   | 0.000| 0.024             | 1.000       | 0.022    |
| 0.05    | 0.000| 0.051             | 1.000       | 0.044    |
| 0.10    | 0.000| 0.104             | 1.000       | 0.093    |
| 0.25    | 0.000| 0.257             | 1.000       | 0.234    |
| 0.50    | 0.000| 0.509             | 1.000       | 0.482    |
| 0.75    | 0.000| 0.755             | 1.000       | 0.736    |
| 0.90    | 0.000| 0.902             | 1.000       | 0.894    |
| 0.95    | 0.001| 0.951             | 1.000       | 0.948    |
| 0.975   | 0.002| 0.974             | 1.000       | 0.972    |
| 0.99    | 0.007| 0.989             | 1.000       | 0.988    |

The integrated likelihood with target weight function, instead, is computed by applying a constant weight function to the likelihood reparameterized with the zero-score-expectation parameter (5), which is $\phi_i = \hat{\psi}_\lambda_i/\psi$. In this case, we obtain an integrated likelihood that is exactly equivalent to the conditional likelihood. It is worth noting that the conditional likelihood for the shape parameter of a gamma is also a marginal likelihood. Hence, the same result can be achieved also using as a weight function $\phi_i^{-1}$, the weight function related with the right invariant measure (see Pace & Salvan, 1997, Example 7.29). As a final remark, we note that the modified profile likelihood in this example is not exactly equivalent to the conditional likelihood (Sartori, 2003, Example 2).

We perform a simulation study with $q = 1000$ and $m = 10$. We chose rather extreme values of $m$ and $q$ in order to investigate the importance of the asymptotic condition on $q/m^3$ in practice. Table 1 shows the empirical coverages of several signed root likelihood ratio statistics based on 8,000 replications, where the nuisance parameters $\lambda_i$ have been generated from.
Table 2: Example 1. Bias and Root Mean Squared Error (RMSE) for different estimators.

|        | \( \hat{\psi} \) | \( \hat{\psi}_C = \hat{\psi}_I \) | \( \hat{\psi}_G \) | \( \hat{\psi}_{MP} \) |
|--------|-------------------|------------------|-----------------|------------------|
| bias   | 0.253             | 0.006            | -1.040          | 0.009            |
| RMSE   | 0.283             | 0.110            | 1.040           | 0.111            |

a \( \chi^2 \) with 10 degrees of freedom and considered fixed in each replication, while the true value of \( \psi \) is 2. The empirical coverages of \( R \) and \( \bar{R}_G \) are, as expected, very poor. Here, with \( \bar{R}_G \) we denote the integrated signed likelihood ratio statistic based on (6), i.e., constructed with a non-target weight function. Using a target weight function in the construction of integrated likelihood, instead, we obtain empirical coverages for the integrated signed likelihood ratio statistics, \( \bar{R}_I \), very close to the nominal ones. Indeed, as seen, it coincides with \( R_C \), the conditional signed likelihood ratio statistic, while the performance of \( R_{MP} \), the modified profile signed likelihood ratio statistic, is overall slightly worse due to the inability of the modified profile log-likelihood to recover \( l_C(\psi) \). The same conclusions can be drawn from bias and root mean squared error of the corresponding estimators, reported in Table 2.

Example 2: matched binomial. Let us consider \( Y_{i1} \) and \( Y_{i2}, i = 1, \ldots, q \), two independent random variables with distribution \( Bi(m, p_{i1}) \) and \( Bi(1, p_{i2}) \) respectively. Let \( \lambda_i = \log\{p_{i1}/(1 - p_{i1})\} \) be the stratum nuisance parameter and \( \psi = \log\{p_{i2}/(1 - p_{i2})\} - \log\{p_{i1}/(1 - p_{i1})\} \) be the parameter of interest, common among strata. We may deal with a model like this in case-control studies where we are interested in studying the effect of a certain factor by
the comparison among one case and \( m \) controls (Sartori, 2003, Example 3).

The likelihood is

\[
L(\psi, \lambda) = \left( \prod_{i=1}^{q} e^{(y_{i1}+y_{i2})\lambda + y_{i2}\psi} \right) / \{(1 + e^\lambda)^m (1 + e^{\psi+\lambda})\}^q,
\]

while the conditional likelihood is a noncentral hypergeometric distribution, see, for instance, Davison (1988, Example 6.1). In order to find a target weight function, we use here an idea suggested by Cox & Reid (1993), i.e., we choose a weight function based on the original parameterization that would act like a uniform one in an orthogonal parameterization, \((\psi, \xi_i)\). Since the model is a full exponential family, \((\psi, \xi_i)\) might be given by the mixed parameterization. Hence, we have \( \partial \xi_i / \partial \lambda_i = me^\lambda_i / (1 + e^\lambda_i)^2 + e^{\psi+\lambda_i} / (1 + e^{\psi+\lambda_i})^2 \), which leads to an integrated likelihood

\[
L_O(\psi) = \prod_{i=1}^{q} \int_{\mathbb{R}} \frac{e^{(y_{i1}+y_{i2})\lambda_i + y_{i2}\psi}}{(1 + e^\lambda_i)^m + 2 (1 + e^{\psi+\lambda_i})} \{e_i^{\lambda_i}(1 + e^{\lambda_i+\psi})^2 + e^{\psi+\lambda_i}(1 + e^\lambda_i)^2\} d\lambda_i.
\]

After a change of variable \( \lambda_i(\omega_i) = \log\{\omega_i/(1 - \omega_i)\} \) and some algebra, we obtain

\[
L_O(\psi) = \prod_{i=1}^{q} e^{y_{i2}\psi} \left\{ 2F_1(1, y_{i1} + y_{i2} + 1, m + 2, 1 - e^\psi) + e^\psi 2F_1(3, y_{i1} + y_{i2} + 1, m + 2, 1 - e^\psi) \right\},
\]

where \( 2F_1(a, b, c, z) = \frac{\Gamma(c)/\{\Gamma(b)\Gamma(c-b)\}}{\int_0^1 x^{b-1}(1-x)^{c-b-1}(1-zx)^{-a} dx} \) (Abramowitz & Stegun, 1964, formula 15.3.1, page 558). We also use the procedure based on the ZSE parameterization (5). Exploiting the exponen-
tial family framework, the new nuisance parameter $\phi_i$ is the solution of the implicit equation

$$K_{\lambda_i}(\hat{\psi}, \phi_i) - K_{\lambda_i}(\psi, \lambda_i) = 0,$$

(8)

where $K$ is the cumulant function and the subscript denotes the derivative with respect to $\lambda_i$. Then we can obtain the integrated likelihood by a change of variable from $\phi_i$ to $\lambda_i$ in the integrals,

$$\bar{L}_I(\psi) = \prod_{i=1}^{q} \int L_i(\psi, \lambda_i) \partial \phi_i(\psi, \lambda_i; \hat{\psi}) \frac{d\phi_i(\psi, \lambda_i)}{d\lambda_i} d\lambda_i$$

$$= \prod_{i=1}^{q} \int L_i(\psi, \lambda_i) \frac{K_{\lambda,\lambda}(\psi, \lambda_i)}{K_{\lambda,\lambda}(\psi, \phi_i(\psi, \lambda_i; \hat{\psi}))} d\lambda_i.$$  

(9)

where the Jacobian $\partial \phi_i(\psi, \lambda_i; \hat{\psi})/\partial \lambda_i$ is obtained by differentiating (8) with respect to $\lambda_i$. Of course we need $\phi_i$ as well in the integrand function; but for fixed $\lambda_i$, $\psi$ and $\hat{\psi}$, it is possible to solve (8) numerically and get the corresponding $\phi_i$.

We perform a simulation study with $q = 300$ and $m = 7$, $\psi = \log(5)$, and $\lambda_i$ equal to 1/8 plus a standard normal random noise. In Table 3 we report the empirical coverage probabilities, based on 8,000 replications, of signed root likelihood ratio statistics based on profile likelihood ($R$), on conditional likelihood ($R_C$), on integrated likelihood with ZSE parameterization and uniform weight function ($\bar{R}_I$), on (7) ($\bar{R}_O$), and on modified profile likelihood ($R_{MP}$). Bias and root mean squared error of the corresponding estimators are reported in Table 4. The empirical coverages of $\bar{R}_I$ and $\bar{R}_O$ are comparable, with the first being slightly more accurate and very close to the behavior of $R_{MP}$. Both $\bar{R}_I$ and $R_{MP}$ give a reasonable approximation for $R_C$ and im-
Table 3: Example 2. Empirical coverage of $R$, $R_C$, $\bar{R}_I$, $\bar{R}_O$, and $R_{MP}$.

| nominal | $R$  | $R_C$ | $\bar{R}_I$ | $\bar{R}_O$ | $R_{MP}$ |
|---------|------|-------|-------------|-------------|----------|
| 0.01    | 0.000| 0.010 | 0.006       | 0.005       | 0.006    |
| 0.025   | 0.000| 0.023 | 0.017       | 0.016       | 0.018    |
| 0.05    | 0.001| 0.050 | 0.036       | 0.033       | 0.037    |
| 0.10    | 0.005| 0.102 | 0.078       | 0.073       | 0.080    |
| 0.25    | 0.026| 0.250 | 0.210       | 0.198       | 0.211    |
| 0.50    | 0.104| 0.497 | 0.448       | 0.429       | 0.448    |
| 0.75    | 0.263| 0.744 | 0.706       | 0.686       | 0.702    |
| 0.90    | 0.471| 0.895 | 0.874       | 0.858       | 0.871    |
| 0.95    | 0.604| 0.946 | 0.935       | 0.925       | 0.932    |
| 0.975   | 0.709| 0.973 | 0.969       | 0.960       | 0.965    |
| 0.99    | 0.815| 0.988 | 0.985       | 0.982       | 0.983    |

Table 4: Example 2. Bias and root mean squared error (RMSE) of different estimators.

|       | $\hat{\psi}$ | $\hat{\psi}_C$ | $\hat{\psi}_I$ | $\hat{\psi}_O$ | $\hat{\psi}_{MP}$ |
|-------|---------------|-----------------|-----------------|-----------------|-------------------|
| bias  | 0.250         | 0.006           | 0.026           | 0.035           | 0.027             |
| RMSE  | 0.320         | 0.171           | 0.170           | 0.176           | 0.174             |

prove substantially over $R$. The close agreement between $\bar{R}_I$ and $R_{MP}$ can be explained by the results for exponential families in Severini (2007). The same indication can be found looking at bias and root mean squared error of the corresponding estimators in Table 4.

**Example 3:** first-order non stationary autoregressive model. Consider the first-order autoregressive model defined by

$$y_{ij} = \lambda_i + \rho y_{ij-1} + \varepsilon_{ij}, \quad (10)$$

where $\varepsilon_{ij}$ are independent normal random variables with zero mean and variance $\sigma^2$, $i = 1, \ldots, q$, $j = 1, \ldots, m$.  

21
When the time series in each stratum are stationary, that is when we assume \( y_{i0} \sim N\{0, \sigma^2/(1 - \rho^2)\} \), then \( \lambda_i \) is orthogonal to \( \psi = (\rho, \sigma^2) \), and the modified profile likelihood and the integrated likelihood proposed in this paper are equivalent and they both coincide with a marginal likelihood. The latter yields consistent estimates for \( \psi \) when \( q \) diverges, even for fixed \( m \) (Bartolucci et al., 2014, Example 1).

Here, we consider the non stationary case, which appears to be the dominant one in the econometric literature (see, for instance, Lancaster, 2002, Section 3). This means that we condition on the observed initial value \( y_{i0} \) and permit the autoregressive parameter to equal or exceed unity. Without loss of generality, in the following we will assume that \( y_{i0} = 0, \ i = 1, \ldots, q \). Indeed, this corresponds to assuming model (10) for the differences \( y_{ij} - y_{i0} \), with \( \lambda_i \) reparameterized as \( \lambda_i - y_{i0}(1 - \rho) \).

The log likelihood for model (10) is the sum of \( q \) independent components of the form

\[
 l_i(\rho, \sigma^2, \lambda_i) = -\frac{m}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^{m} (y_{ij} - \lambda_i - \rho y_{i,j-1})^2 .
\]

Lancaster (2002, page 655) shows that an information orthogonal parameterization is given by \( \xi_i = \lambda_i \exp\{b(\rho)\} \), where

\[
 b(\rho) = \frac{1}{m} \sum_{j=1}^{m-1} \frac{m-j}{j} \rho^j .
\]  (11)

The parameter \( \xi_i \) is orthogonal to both \( \rho \) and \( \sigma^2 \), with the latter two being orthogonal to each other.
Alternatively, we can use the ZSE parameterization (5), with \( \phi_i \) solution of
\[
E_{\rho_0,\sigma^2,\lambda_0} \{ l_\lambda_i(\rho,\sigma^2,\lambda_i) \} | (\rho_0,\sigma^2,\lambda_0) = (\hat{\rho},\hat{\sigma}^2,\phi_i) = 0 .
\] (12)

Using again the results in Lancaster (2002), we find \( \phi_i = \lambda_i / \{1 + (\hat{\rho} - \rho)b'(\hat{\rho})\} \), where \( b'(\rho) = (1/m) \sum_{j=1}^{m-1} (m-j)\rho^{j-1} \) is the first derivative of (11). For this model computation of profile and integrated log likelihoods is straightforward since all maximization and integration involved can be easily done analytically. In particular, focusing interest on the parameter \( \rho \), we have
\[
\begin{align*}
l_p(\rho) &= -\frac{mq^2}{2} \log SS(\rho) , \\
l_o(\rho) &= -\frac{m(q-1)}{2} \log SS(\rho) + qb(\rho) , \\
l_i(\rho) &= -\frac{m(q-1)}{2} \log SS(\rho) - q \log \{1 + (\hat{\rho} - \rho)b'(\hat{\rho})\} ,
\end{align*}
\] (13) (14)
where \( SS(\rho) = \sum_{i=1}^{q} \sum_{j=1}^{m} \{w_{ij}(\rho) - \bar{w}_i(\rho)\}^2 \), with \( w_{ij}(\rho) = y_{ij} - \rho y_{ij-1} \), and \( \bar{w}_i(\rho) = m^{-1} \sum_{j=1}^{m} w_{ij}(\rho) \). Formulae (13) and (14) are the integrated log-likelihoods with the orthogonal parameters \( \xi_i \) and with the ZSE parameters \( \phi_i \), respectively. In both cases we used a constant weight function for the incidental parameters and for \( \log \sigma \). These integrated log likelihoods could be also obtained by first integrating out the incidental parameters, thus obtaining the integrated log likelihoods for \((\rho,\sigma^2)\), and then profiling out \( \sigma^2 \).

Since the maximum likelihood estimate is generally highly biased, this could have an effect on the accuracy of the integrated likelihood (14). A possible solution could be given by using in (12) alternative estimates for \( \rho \).
and $\sigma^2$ in place of $\hat{\rho}$ and $\hat{\sigma}^2$. One solution could be the use of a parametric bootstrap bias corrected version of $\hat{\rho}$ and $\hat{\sigma}^2$. Alternatively, one could use a different estimate, such as for instance the maximizer of (13), or the maximizer of (14) itself, leading to a two-step solution. In the numerical example and in the simulations below we used the former option, thus obtaining the new ZSE parameter $\phi^I_i = \lambda_i / \{1 + (\hat{\rho}_O - \rho) b'(\hat{\rho}_O)\}$, where $\hat{\rho}_O$ denote the maximizer of (13). The corresponding integrated likelihood has the form (14), with $\hat{\rho}$ replaced by $\hat{\rho}_O$, and will be denoted by $\tilde{l}_I(\rho)$.

Sometimes $l_O(\rho)$ can be monotonic increasing for large values of $\rho$. On the other hand, for values of $m$, $q$ and $\rho$ of practical interest, it has a local maximum for $\rho \in (-\rho_l, \rho_u)$, where $\rho_l, \rho_u > 0$ are threshold values that can exceed one. Lancaster (2002), developing the integrated likelihood from a Bayesian perspective, shows that such a local maximum is a consistent estimator of $\rho$ for large $q$, even for fixed $m$. Also $\tilde{l}_I(\rho)$ and $\hat{l}_I(\rho)$ can be monotonic increasing for large values of $\rho$, and this problem seems to occur “sooner” than for $l_O(\rho)$. Moreover, the second term in the right hand side of (14) cannot be computed for values of $\rho$ greater than $\hat{\rho} + 1 / b'(\hat{\rho})$ (which is however always greater than 1). A similar comment applies also to $\tilde{l}_I(\rho)$. Even in these cases, in practice, this has not proven to be a problem for maximization and inference.

As a numerical illustration, Figure 1 shows the relative log likelihoods for a simulated sample with $m = 8$, $q = 500$, $\rho = 0.9$, $\sigma^2 = 1$ and $\lambda_i$ generated from a normal with mean and variance equal to 1. The left panel shows the monotonicity issue for the integrated log likelihoods, while the right panel gives a zoomed version in an interval of values of practical interest for inference.
Figure 1: Example 3. Relative log likelihoods for simulated data of the nonstationary autoregressive model: $m = 8$, $q = 500$, $\rho = 0.9$, $\sigma^2 = 1$ and $\lambda_i \sim N(1, 1)$. The solid line corresponds to $l_P(\rho)$, the dashed line to $l_O(\rho)$, the dot-dashed line to $\bar{l}_I(\rho)$, and the long-dashed line to $\tilde{l}_I(\rho)$. The vertical dotted line indicates the true parameter value, while the horizontal dotted line provides confidence intervals of level 0.95 based on the corresponding likelihood ratio statistics. The left panel shows the unconstrained plot, while the right panel shows a zoomed version in a region of interest.
We also run some simulation studies comparing the empirical coverage probabilities for the signed likelihood ratio statistics based on \( l_P(\rho) \), \( l_O(\rho) \), \( \bar{l}_I(\rho) \), and \( \tilde{l}_I(\rho) \), which are denoted by \( R \), \( R_O \), \( \bar{R}_I \) and \( \tilde{R}_I \), respectively. Also bias and mean squared errors of the corresponding estimators have been considered. Tables 5 and 6 report the results for the same setting of Figure 1, with 10,000 simulated samples. The results indicate that \( \bar{l}_I(\rho) \), although largely improving over \( l_P(\rho) \), is far from having the same accuracy of \( l_O(\rho) \). On the other hand, \( \tilde{R}_I \) seems to be slightly more accurate than \( R_O \), in particular in the tails. Results in more extreme settings, such as with \( m = 4 \) and \( q = 1000 \), confirm these findings.

| nominal | \( R \) | \( R_O \) | \( \bar{R}_I \) | \( \tilde{R}_I \) |
|---------|--------|--------|--------|--------|
| 0.010   | 1.000  | 0.014  | 0.003  | 0.013  |
| 0.025   | 1.000  | 0.031  | 0.011  | 0.027  |
| 0.050   | 1.000  | 0.056  | 0.021  | 0.050  |
| 0.100   | 1.000  | 0.105  | 0.041  | 0.100  |
| 0.250   | 1.000  | 0.259  | 0.122  | 0.252  |
| 0.500   | 1.000  | 0.496  | 0.312  | 0.496  |
| 0.750   | 1.000  | 0.746  | 0.568  | 0.753  |
| 0.900   | 1.000  | 0.896  | 0.781  | 0.903  |
| 0.950   | 1.000  | 0.944  | 0.872  | 0.949  |
| 0.975   | 1.000  | 0.971  | 0.925  | 0.975  |
| 0.990   | 1.000  | 0.988  | 0.963  | 0.990  |

| \( \hat{\rho} \) | \( \hat{\rho}_O \) | \( \hat{\rho}_I \) | \( \hat{\rho}_{\tilde{I}} \) |
|-----------------|-----------------|-----------------|-----------------|
| bias            | -0.052          | 1.0 \times 10^{-3} | 3.38 \times 10^{-3} | -1.2 \times 10^{-3} |
| RMSE            | 0.0525          | 0.0067          | 0.0078          | 0.0067          |

| \( \hat{\sigma}^2 \) | \( \hat{\sigma}_O^2 \) | \( \hat{\sigma}_I^2 \) | \( \hat{\sigma}_{\tilde{I}}^2 \) |
|-----------------|-----------------|-----------------|-----------------|
| bias            | -0.144          | -3.06 \times 10^{-4} | 2.58 \times 10^{-3} | -3.16 \times 10^{-4} |
| RMSE            | 0.1455          | 0.0244          | 0.0248          | 0.0244          |
Finally, we note that the modified profile likelihood is not straightforward to obtain in this model. A possibility is to use the approximation of Severini (1998), avoiding the sometimes cumbersome analytical calculation of required expected value by means of Monte Carlo simulation. This approach is quite general, although computationally more intensive, and has been used also by Bartolucci et al. (2014) for a dynamic regression model for binary data. Claudia Di Caterina, in an unpublished Master Thesis of the University of Padova, proved that Severini’s approximation of the mixed derivative is linear in $\rho$. This implies that the modified profile log likelihood does not exist for certain values of $\rho$, similarly to $\bar{l}_I(\rho)$. Moreover, it also shares the other drawbacks of $\bar{l}_I(\rho)$, i.e., it could be monotonic increasing for not very large values of $\rho$, and the normal approximation for the corresponding signed likelihood ratio statistic has an accuracy very close to that of $\bar{R}_I$, which is not very satisfactory for practical purposes when $m$ is moderate and $q$ is very large. We note that also the modified profile likelihood depends on the maximum likelihood estimates. Therefore it is likely that the use of better estimates could improve also its accuracy, as for the integrated likelihood $\bar{l}_I(\rho)$, although, to our knowledge, this has not been investigated yet.

References

Abramowitz, M. & Stegun, I. A. (1964). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications, New York.

Bahadur, R. (1964). On Fisher’s bound for asymptotic variances. Annals
Barndorff-Nielsen, O. E. (1996). Two order asymptotic. In Frontiers in Pure and Applied Probability II: Proceedings of the Fourth Russian-Finnish Symposium Prob. Th. Math. Statist., A. Melnikov, ed. TVP Science, Moscow.

Barndorff-Nielsen, O. E. & Cox, D. R. (1994). Inference and Asymptotics. London: Chapman and Hall.

Bartolucci, F., Bellio, R., Salvan, A. & Sartori, N. (2014). Modified profile likelihood for fixed effects panel data models. Econometric Reviews, to appear.

Bellio, R. & Guolo, A. (2014). Integrated likelihood inference in small sample meta-analysis. Under second revision.

Berger, J. O., Liseo, B. & Wolpert, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters. Statistical Science 14, 1–22.

Cox, D. R. & Reid, N. (1987). Parameter orthogonality and approximate conditional inference. J. Roy. Statist. Soc. B 49, 1–39.

Cox, D. R. & Reid, N. (1993). A note on the calculation of adjusted profile likelihood. J. Roy. Statist. Soc. B 55, 467–471.

Davison, A. C. (1988). Approximate conditional inference in generalized linear models. J. Roy. Statist. Soc. B 50, 445–461.

Efron, B. (1993). Bayes and likelihood calculations from confidence intervals. Biometrika 80, 3–26.
Kalbfleisch, J. D. & Sprott, D. A. (1970). Application of likelihood methods to models involving large numbers of parameters. *J. Roy. Statist. Soc. B* 32, 175–208.

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

Pace, L. & Salvan, A. (1997). *Principles of Statistical Inference: from a neo-Fisherian perspective*. World Scientific, Singapore.

Pace, L. & Salvan, A. (2006). Adjustments of the profile likelihood from a new perspective. *Journal of Statistical Planning and Inference* 136, 3554–3564.

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

Severini, T. A. (1998). An approximation to the modified profile likelihood function. *Biometrika* 85, 403–411.

Severini, T. A. (2000). *Likelihood Methods in Statistics*. Oxford University Press, Oxford.

Severini, T. A. (2007). Integrated likelihood functions for non-Bayesian inference. *Biometrika* 94, 529–542.

Severini, T. A. (2010). Likelihood ratio statistics based on an integrated likelihood. *Biometrika* 97, 481–496.

Severini, T. A. (2011). Frequency properties of inferences based on an integrated likelihood function. *Statistica Sinica* 21, 433.
VENTURA, L., CABRAS, S. & RACUGNO, W. (2009). Prior Distributions From Pseudo-Likelihoods in the Presence of Nuisance Parameters. *Journal of the American Statistical Association* **104**, 768–774.