Simulation based Bias Correction Methods for Complex Models

Stéphane Guerrier[§], Elise Dupuis-Lozeron[†], Yanyuan Ma[§] &
Maria-Pia Victoria-Feser[†]

[§]Department of Statistics
Pennsylvania State University
University Park, PA 16802, USA

[†]Research Center for Statistics
Geneva School of Economics and Management
University of Geneva
Geneva, Switzerland

Abstract

Along the ever increasing data size and model complexity, an important challenge frequently encountered in constructing new estimators or in implementing a classical one such as the maximum likelihood estimator, is the computational aspect of the estimation procedure. To carry out estimation, approximate methods such as pseudo-likelihood functions or approximated estimating equations are increasingly used in practice as these methods are typically easier to implement numerically although they can lead to inconsistent and/or biased estimators. In this context, we extend and provide refinements on the known bias

correction properties of two simulation based methods, respectively indirect inference and bootstrap, each with two alternatives. These results allow one to build a framework defining simulation based estimators that can be implemented for complex models. Indeed, based on a biased or even inconsistent estimator, several simulation based methods can be used to define new estimators that are both consistent and with reduced finite sample bias. This framework includes the classical method of indirect inference for bias correction without requiring specification of an auxiliary model. We demonstrate the equivalence between one version of the indirect inference and the iterative bootstrap, both correct sample biases up to the order n^{-3}. The iterative method can be thought of as a computationally efficient algorithm to solve the optimization problem of the indirect inference. Our results provide different tools to correct the asymptotic as well as finite sample biases of estimators and give insight on which method should be applied for the problem at hand. The usefulness of the proposed approach is illustrated with the estimation of robust income distributions and generalized linear latent variable models.

Keywords: *Iterative bootstrap; Two-step estimators; Indirect inference; Robust statistics; Weighted maximum likelihood estimators; Generalized latent variable models.*

1 Introduction

Estimation of parameters of complex statistical models often leads to difficult numerical problems. This is a difficulty shared by the classical Maximum Likelihood Estimator (MLE) and other estimators. For example, in the finance domain, stochastic models representing different unobserved processes are proposed on a continuous timeline while only observations reflecting the joint effect of these processes are available, usually on discrete times. Latent variable models, such as state space models (see e.g. Harvey, 1990 and Durbin and Koopman, 2001) or Generalized Linear Latent Variable Models (GLLVM) (see e.g. Bartholomew, 1984 and Moustaki and Knott, 2000) are such typical models and the corresponding estimation procedure often implies integrating out from the likelihood equations the unobserved variables, leading to multiple integrals that have no analytical solution. Another example is in robust estimation, where the consistency correction in an estimating equation often renders the computation very difficult or nearly infeasible while the estimating equation without consistency correction is numerically relatively simple.

To overcome these numerical challenges, many strategies have been used, and they are usually different from model to model, or application to application. For intractable integrals, approximations such as adaptive quadratures or Laplace approximations can be used. In time series, stochastic models are often discretized to yield an approximate model and then classical estimators

are used which lead to biased estimates (see e.g. Gouriéroux and Monfort, 1997). In such cases, simulation methods are used to correct the biases of the estimators. In robust statistics, to ease the computational burden, two-step estimators can be built that analytically correct the linear part of the first step estimator's bias (see e.g. Dupuis and Morgenthaler, 2002). With latent variable models, alternative target functions such as composite likelihoods (see Lindsay, 1988) are used to replace the likelihood function. Strategies used to overcome computational difficulties in estimating complex models are mostly tailor made to specific (classes of) models and their application outside the models they are used for is, in general, not straightforward, and sometimes even impossible.

In this paper, we study the properties of several simulation based methods which can be used to correct sample and asymptotic biases. The resulting properties allow one to build a framework defining simulation based estimators that can be implemented very generally for complex models. This framework includes for example the method of indirect inference for bias correction, but does not necessarily rely on the specification of an auxiliary model. The estimators considered in this paper are two-steps estimators, in which the first step proposes an estimator that is easy to compute, but is not necessarily consistent, while the second step corrects biases via simulations. In practice, the relatively small price to pay is a sample variance increase (of a quantifiable order) of the simulation based estimator relative to the sample variance of a direct consistent estimator, if computable without numerical

approximations. As a by product, some of the simulation based estimators can have a reduced sample bias of order up to n^{-3}.

More specifically, we consider four existing simulation methods, two of them being based on the indirect inference method (Gouriéroux et al., 1993, Smith, 1993 and Gallant and Tauchen, 1996a) and two on the boostrap. The bootstrap is mainly used to reduce sample bias (Efron and Tibshirani, 1993), but Kuk (1995) propose an iterative bootstrap for asymptotic bias correction, that is for example used in Mealli and Rampichini (1999) for Generalized Linear Mixed Models (GLMM). Indirect inference is mainly used to correct for asymptotic bias caused by the necessary approximations induced by the specification of an auxiliary model. In this context, Arvanitis and Demos (2015) define a set of indirect inference estimators based on moment approximations of the auxiliary estimators for essentially analytical and computational facilitation. Gouriéroux et al. (2000) have studied the sample bias correction properties of indirect inference. Arvanitis and Demos (2013) (see also Gallant and Tauchen, 1996b, Arvanitis and Demos, 2014) define, within the framework of indirect inference, classes of estimators based on analytical approximations of either the binding or the score function of the auxiliary estimator. They study and compare their properties on the resulting bias convergence order. In this paper, the emphasis is put on estimators for very complex models and/or complex estimators like robust estimators. We therefore concentrate on the classical indirect inference (without analytical bias correction) and the bootstrap, and find that the iterative bootstrap is equiv-

alent to (one version of) indirect inference. This hence provides an easy to implement algorithm to obtain the indirect inference estimator.

In Section 2 we formally set up two versions of indirect inference, the bootstrap and the iterative bootstrap. The two versions of indirect inference differ in the simulation of either H samples of size n or one sample of size nH, and the conventional bootstrap corresponds to the first step of the iterative bootstrap. In Section 3 we study the properties of the four resulting estimators, and find out that the indirect inference with H samples of size n is equivalent to the iterative bootstrap, both providing a higher order reduction in terms of bias (and often variance) over the other two. Basically, a higher order bias reduction is achieved when using the version of indirect inference with H samples of size n over the one sample of size nH, and the same is true for the iterative bootstrap over conventional bootstrap. Gouriéroux et al. (2000) studied the finite sample bias correction properties of indirect inference procedures via Edgeworth expansions. Our results provide an alternative study, which in some cases, lead to bias correction of higher order. Also, MacKinnon and Smith (1998) study the relationship between the nature of the bias and the method that should be used to reduce the sample bias more effectively. We show instead that for a large class of bias structures, iterative bootstrap and indirect inference achieve overall sample bias reduction of order n^{-3}. We also demonstrate the equivalence between one version of the indirect inference and the iterative bootstrap, hence providing iterative bootstrap as a computationally efficient algorithm to solve the

optimization problem involved in the indirect inference procedure.

As an application, we propose new simulation based estimators in two different settings, that are drastically easier to compute especially in complex models. The first one provided in Section 6 is for robust estimation in general, and for robust income distribution estimation in particular, and the second one provided in Section 7 is for estimation of GLLVM.

2 Mathematical Setup

Let the true parameter value $\boldsymbol{\theta}_0$ be an interior point of a compact convex set $\boldsymbol{\Theta} \subset \mathbb{R}^p$. Let $\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n)$ be an estimator of $\boldsymbol{\theta}_0$ based on a sample of size n generated from $F_{\boldsymbol{\theta}_0}$. We assume that this estimator has a non-stochastic limit in the sense that

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} ||\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}, n) - \boldsymbol{\pi}(\boldsymbol{\theta})||_2 \xrightarrow{p} 0. \tag{1}$$

Typically $\boldsymbol{\pi}(\boldsymbol{\theta}_0) \neq \boldsymbol{\theta}_0$, hence $\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n)$ is not a consistent estimator of $\boldsymbol{\theta}_0$ but is readily available (e.g. easier to compute). Assume $\mathbb{E}[\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n)]$ exists, where $\mathbb{E}[\cdot]$ denotes the expectation under $F_{\boldsymbol{\theta}_0}$. We let the bias be $\mathbf{d}(\boldsymbol{\theta}_0, n) \equiv \mathbb{E}[\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n)] - \boldsymbol{\theta}_0$. We can write

$$\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n) = \boldsymbol{\pi}(\boldsymbol{\theta}_0, n) + \mathbf{v}(\boldsymbol{\theta}_0, n), \tag{2}$$

where $\boldsymbol{\pi}(\boldsymbol{\theta}_0, n) \equiv \mathbb{E}[\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n)] = \boldsymbol{\theta}_0 + \mathbf{d}(\boldsymbol{\theta}_0, n)$ and $\mathbf{v}(\boldsymbol{\theta}_0, n) \equiv \hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n) - \mathbb{E}[\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n)]$ is a zero mean random vector.

We now describe four bias correction strategies. The first two are indirect estimators in the sense of Gouriéroux et al. (1993). The third strategy is the first-order bootstrap correction of Efron and Tibshirani (1993), and the last strategy is an iterative version of the third strategy proposed in Kuk (1995).

STRATEGY 1: We define indirect estimator based on H samples of size n as

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argzero}} \ \hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n) - \bar{\boldsymbol{\pi}}(\boldsymbol{\theta}, n), \tag{3}$$

where

$$\bar{\boldsymbol{\pi}}(\boldsymbol{\theta}, n) = \frac{1}{H} \sum_{h=1}^{H} \hat{\boldsymbol{\pi}}_h(\boldsymbol{\theta}, n), \tag{4}$$

where $\hat{\boldsymbol{\pi}}_h(\boldsymbol{\theta}, n)$ denotes the value of $\hat{\boldsymbol{\pi}}$ obtained from the h^{th} simulated sample of size n under $F_{\boldsymbol{\theta}}$.

STRATEGY 2: We define the indirect estimator based on one sample of size nH as

$$\tilde{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argzero}} \ \hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n) - \hat{\boldsymbol{\pi}}(\boldsymbol{\theta}, nH), \tag{5}$$

where $\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}, nH)$ denotes the value of $\hat{\boldsymbol{\pi}}$ obtained from one simulated sample of size nH under $F_{\boldsymbol{\theta}}$.

REMARK A: *The definitions used in Strategies 1 and 2 for indirect estimators*

are not standard. Indeed, such estimators are typically defined as

$$\tilde{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \; ||\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n) - \hat{\boldsymbol{\pi}}^*(\boldsymbol{\theta})||^2_{\boldsymbol{\Phi}}, \tag{6}$$

where $\boldsymbol{\Phi}$ *is a positive-definite matrix and* $\hat{\boldsymbol{\pi}}^*(\boldsymbol{\theta})$ *is a suitable estimator of* $\boldsymbol{\pi}(\boldsymbol{\theta})$ *or* $\boldsymbol{\pi}(\boldsymbol{\theta}, n)$*. In our context,* $\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}, n)$ *is an estimator of* $\boldsymbol{\theta}_0$*, implying that* $dim(\boldsymbol{\pi}) = dim(\boldsymbol{\theta})$*. Therefore, the minimization problem defined in (6) can be expressed as the solution in* $\boldsymbol{\theta}$ *of* $\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n) = \hat{\boldsymbol{\pi}}^*(\boldsymbol{\theta})$ *or using the* argzero *operator as in Strategies 1 and 2, provided that the minimum of the quadratic form defined in (6) is attained in the interior of* $\boldsymbol{\Theta}$*.*

STRATEGY 3: The bootstrap bias corrected estimator is defined as:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_B &= \hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n) + \left[\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n) - \bar{\boldsymbol{\pi}}\left\{\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n), n\right\}\right], \\ &= 2\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n) - \bar{\boldsymbol{\pi}}\left\{\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n), n\right\}, \end{aligned} \tag{7}$$

where $\bar{\boldsymbol{\pi}}\{\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n), n\}$ is calculated as in (4) but with $\boldsymbol{\theta}$ replaced by $\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n)$. Here the bias $\boldsymbol{\theta}_0 - \mathbb{E}[\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n)]$ is assessed through sampling H samples of size n from $F_{\hat{\boldsymbol{\pi}}}$.

STRATEGY 4: At iteration k, the iterative bootstrap bias corrected estimator $\tilde{\boldsymbol{\theta}}_B^{(k)}$ is defined as:

$$\tilde{\boldsymbol{\theta}}_B^{(k)} = \hat{\boldsymbol{\pi}}(\boldsymbol{\theta}_0, n) + \left\{\tilde{\boldsymbol{\theta}}_B^{(k-1)} - \bar{\boldsymbol{\pi}}\left(\tilde{\boldsymbol{\theta}}_B^{(k-1)}, n\right)\right\}, \tag{8}$$

with

$$\bar{\boldsymbol{\pi}}\left(\tilde{\boldsymbol{\theta}}_B^{(k-1)}, n\right) = \frac{1}{H}\sum_{h=1}^{H}\hat{\boldsymbol{\pi}}_h\left(\tilde{\boldsymbol{\theta}}_B^{(k-1)}, n\right).$$

We define $\tilde{\boldsymbol{\theta}}_B$ as the limit of $\tilde{\boldsymbol{\theta}}_B^{(k)}$ once the procedure converges. In the proof of Theorem 4 given in the supplementary material E, we prove that the limit of $\tilde{\boldsymbol{\theta}}_B^{(k)}$ indeed exists (under some appropriate conditions).

REMARK B: *In all the strategies, we assume that synthetic samples can be generated under the model $F_{\boldsymbol{\theta}}$, i.e. the model is parametric. In most of the cases where bias correction techniques are employed, $F_{\boldsymbol{\theta}}$ fully defines the data generating process. In some cases, it is possible to extend the parametric model setting to the semi-parametric case, while the extension to a general semi-parametric model can be challenging and is worth further research.*

Before presenting the statistical properties of the four bias correction methods, we first describe the assumptions we use.

ASSUMPTION 1: *The function $\boldsymbol{\pi}(\boldsymbol{\theta})$ is continuous and one-to-one on $\boldsymbol{\Theta}$.*

ASSUMPTION 2: *Let $n^{-\alpha}\mathbf{V}_{\theta,n}$ be the variance-covariance matrix of $\mathbf{v}(\boldsymbol{\theta}, n)$ given in (2). Then $\alpha > 0$, and $||\mathbf{V}_{\boldsymbol{\theta},n}||_1 < \infty$ for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and at any sample size n.*

Assumption 1 ensures identifiability. Assumption 2 requires that the variance of $\mathbf{v}(\boldsymbol{\theta}, n)$ goes to zero as n increases. For example, for many estimators, $\alpha = 1$. Assumptions 1 and 2, are quite mild and frequently employed. These assumptions directly lead to Corollary 1, whose proof is based on Theorem

2.1 of Newey and McFadden (1994) and is given in Appendix A in the Supplementary Materials.

COROLLARY 1: *Under the Assumptions 1 and 2, $\hat{\boldsymbol{\theta}}$ defined in (3), $\tilde{\boldsymbol{\theta}}$ defined in (5) and $\tilde{\boldsymbol{\theta}}_B$ defined after (8) are consistent estimators of $\boldsymbol{\theta}$.*

Corollary 1 only considers Strategies 1, 2 and 4 since the classical one-step bootstrap (i.e. Strategy 3) does not necessarily provide a consistent estimator under the general conditions considered here. Next, we impose some additional assumptions on the form of the bias vector $\mathbf{d}(\boldsymbol{\theta}_0, n)$.

ASSUMPTION 3: *The bias vector $\mathbf{d}(\boldsymbol{\theta}, n)$ can be expressed as follows:*

$$\mathbf{d}(\boldsymbol{\theta}, n) = \mathbf{c}(n) + \mathbf{b}(\boldsymbol{\theta}, n), \tag{9}$$

and $\exists a_{i,j} \in \mathbb{R}, d_{k,l,j} \in \mathbb{R}, i, j, k, l = 1, \ldots, p$ such that

$$\mathbf{b}(\boldsymbol{\theta}, n) = \left[\sum_{i=1}^{p} a_{i,j} \frac{\theta_i}{n} + \sum_{k=1}^{p} \sum_{l=1}^{p} d_{k,l,j} \frac{\theta_k \theta_l}{n^2} + \mathcal{O}\left(n^{-3}\right) \right]_{j=1,\ldots,p}. \tag{10}$$

Moreover, $\mathbf{c}(n)$ is bounded, converges to a constant $\mathbf{c}$ when $n \to \infty$ and

$$\mathbf{c}(n) = \mathcal{O}\left(n^{-\beta}\right), \quad \beta \geq 0. \tag{11}$$

Here and throughout the text, for a vector or a matrix $\mathbf{a}$, $\mathbf{a} = \mathcal{O}(n^{-\alpha})$ implies each component of $\mathbf{a}$ is of order $\mathcal{O}(n^{-\alpha})$. Assumption 3 essentially employees a Taylor expansion of $\mathbf{b}(\boldsymbol{\theta}, n)$ as a function of $\boldsymbol{\theta}/n$, and is generally

always applicable when $\mathbf{b}(\boldsymbol{\theta}, n)$ is a sufficiently smooth function of $\boldsymbol{\theta}/n$. Similar approximations are commonly used to assess the bias of simulation-based methods (see e.g. Efron and Tibshirani, 1993).

REMARK C: *A direct consequence of Assumption 3 is that* $\mathbf{c}(n) - \mathbf{c}(Hn) = \mathcal{O}(n^{-\lambda})$, *componentwise and where* $\lambda \geq \beta \geq 0$.

The next assumption imposes some continuity restrictions (with respect to $\boldsymbol{\theta}$) on the matrix $\mathbf{V}_{\boldsymbol{\theta},n}$ defined in Assumption 2.

ASSUMPTION 4: *Let* $\mathbf{w}_j(\boldsymbol{\theta}, n)$ *denote the* j^{th} *column of the matrix* $\mathbf{V}_{\boldsymbol{\theta},n}$, *then we have that for* $j = 1, \ldots, p$ *and any sample size* n, *the matrix*

$$\mathbf{D}_j(\boldsymbol{\theta}, n) \equiv \frac{\partial}{\partial \boldsymbol{\theta}^T} \mathbf{w}_j(\boldsymbol{\theta}, n),$$

exists and is such that $||\mathbf{D}_j(\boldsymbol{\theta}, n)||_1 < \infty$ *for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Assumptions 1 to 4 are very mild and likely satisfied in most practical situations including the examples discussed in Sections 4 to 7. However, they are not necessarily the weakest possible in theory and may be further relaxed. We however do not attempt to pursue the weakest possible conditions to avoid overly technical treatments in establishing the theoretical results in Section 3.

3 Statistical Properties of Simulation Based Methods for Bias Reduction

We now study the convergence rate of the four bias reduction methods presented in Section 2. In Theorem 1, we show that indirect inference provides consistent estimators with a fast convergence rate for any H. Moreover, the variance of the indirect estimator is only slightly inflated compared to the variance of the initial estimator, with an increment of order $\mathcal{O}\left(n^{-\min(2,1+\alpha)}\right)$. The proof of Theorem 1 is given in Appendix B (in the Supplementary Materials).

THEOREM 1: *Under Assumptions 1 to 4, the corrected estimator $\hat{\boldsymbol{\theta}}$ defined in (3) satisfies*

$$\begin{aligned}\mathbb{E}\left[\hat{\boldsymbol{\theta}}\right] &= \boldsymbol{\theta}_0 + \mathcal{O}\left(n^{-\min(3,\alpha+2)}\right), \\ \text{var}\left(\hat{\boldsymbol{\theta}}\right) &= n^{-\alpha}\left(1+\frac{1}{H}\right)\mathbf{V}_{\boldsymbol{\theta}_0,n} + \mathcal{O}\left(n^{-\min(2,1+\alpha)}\right).\end{aligned}$$

An important consequence of Theorem 1 is that the corrected estimator $\hat{\boldsymbol{\theta}}$ operates at least a second order bias correction regardless of the number of simulations H. On the other hand, the variance of the estimator is clearly dependent on H. Therefore when $\alpha = 1$ and H is sufficiently large, the

corrected estimator $\hat{\boldsymbol{\theta}}$ is such that

$$\mathbb{E}\left[\hat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta}_0 + \mathcal{O}\left(n^{-3}\right),$$
$$\text{var}\left(\hat{\boldsymbol{\theta}}\right) \approx n^{-1}\mathbf{V}_{\boldsymbol{\theta}_0,n} + \mathcal{O}\left(n^{-2}\right).$$

In Theorem 2, we show that for the indirect inference procedure, using a single sample of size nH instead of H samples of size n, the asymptotic properties differ. In particular, the sample bias of the single-sample based estimator $\tilde{\boldsymbol{\theta}}$ is larger than that of the multi-sample based estimator $\hat{\boldsymbol{\theta}}$ and depends on the number of simulations H. The proof of Theorem 2 is given in Appendix C (in the Supplementary Materials).

THEOREM 2: *Under Assumptions 1 to 4, the corrected estimator $\tilde{\boldsymbol{\theta}}$ defined in (5) satisfies for $H > 1$*

$$\mathbb{E}\left[\tilde{\boldsymbol{\theta}}\right] = \boldsymbol{\theta}_0 + \mathcal{O}\left(n^{-\min(2,\lambda)}\right),$$
$$\text{var}\left(\tilde{\boldsymbol{\theta}}\right) = n^{-\alpha}\mathbf{V}_{\boldsymbol{\theta}_0,n} + (nH)^{-\alpha}\mathbf{V}_{\boldsymbol{\theta}_0,nH} + \mathcal{O}\left(\max\left\{n^{-\min(2,\lambda)}, H^{-1}n^{-\frac{\min(\alpha,\lambda)}{2}-1}\right\}\right),$$

with λ defined in Remark C.

In order to compare the two versions of indirect inference in a simple setting we consider $\alpha = 1$ and assume that H is sufficiently large, then the corrected estimator $\tilde{\boldsymbol{\theta}}$ is such that

$$\mathbb{E}\left[\tilde{\boldsymbol{\theta}}\right] \approx \boldsymbol{\theta}_0 + \mathcal{O}\left(n^{-\min(2,\lambda)}\right),$$
$$\text{var}\left(\tilde{\boldsymbol{\theta}}\right) \approx n^{-1}\mathbf{V}_{\boldsymbol{\theta}_0,n} + \mathcal{O}\left(n^{-\min(2,\lambda)}\right).$$

Therefore, in this simplified setting the indirect inference procedure based on a single sample of size nH operates at best a second order correction while indirect inference based on H samples of size n achieves a third order correction. Regarding the variance of the estimators, the indirect inference procedure based on a single sample has a remainder that is at best equivalent to the one of the procedure based on multiple samples.

REMARK D: *The results of Theorems 1 and 2 are implicitly based on the assumption that the root defining the estimators can be solved exactly. However, in practice, numerical error always occurs hence the zero (or minimizer) found from an optimizer is almost never the exact solution. Nevertheless, the validity of Theorems 1 and 2 can be extended to this practical situation by assuming that the distance between the solution found by the optimizer and the true solution is of higher order than the ones of the leading bias and standard deviation.*

In Theorem 3, we derive the properties of the one-step bootstrap estimator $\hat{\boldsymbol{\theta}}_B$. The proof is given in Appendix D (in the Supplementary Materials).

THEOREM 3: *Assumptions 1 to 4, the corrected estimator $\hat{\boldsymbol{\theta}}_B$ defined in (7) satisfies*

$$\begin{aligned}
\mathbb{E}\left[\hat{\boldsymbol{\theta}}_B\right] &= \boldsymbol{\theta}_0 + \mathcal{O}\left(n^{-\min(1,\beta)-1}\right), \\
\operatorname{var}\left(\hat{\boldsymbol{\theta}}_B\right) &= n^{-\alpha}\left(1+\frac{1}{H}\right)\mathbf{V}_{\boldsymbol{\theta}_0,n} \\
&\quad + \mathcal{O}\left(\max\left\{n^{-\frac{\min(2,\alpha)}{2}-1}, H^{-1}n^{-\min(1,\beta)}, H^{-1/2}n^{-\frac{\min(1,\beta)}{2}-1}\right\}\right),
\end{aligned}$$

Theorem 3 illustrates that the one-step bootstrap estimator achieves at best a second order correction and can only perform a first order correction when $\beta = 0$. With $\alpha = 1$, the most common case, and with sufficiently large H, this estimator is such that

$$\mathbb{E}\left[\hat{\boldsymbol{\theta}}_B\right] \approx \boldsymbol{\theta}_0 + \mathcal{O}\left(n^{-\min(1,\beta)-1}\right),$$
$$\text{var}\left(\hat{\boldsymbol{\theta}}_B\right) \approx n^{-1}\mathbf{V}_{\boldsymbol{\theta}_0,n} + \mathcal{O}\left(n^{-\frac{\min(1,\beta)}{2}-1}\right).$$

Therefore, in this simplified setting we can expect the one-step bootstrap estimator to achieve a larger bias correction than the indirect inference procedure based on a single sample but a smaller than the indirect inference approach based on multiple samples.

Finally, Theorem 4 proves the equivalence of the iterative bootstrap bias correction estimator with the multiple sample indirect inference estimator $\hat{\boldsymbol{\theta}}$. The proof is given in Appendix E (in the Supplementary Materials).

THEOREM 4: *Under Assumption 3, for sufficiently large n and H, the iterative bootstrap is equivalent to $\hat{\boldsymbol{\theta}}$ in the sense that*

$$\tilde{\boldsymbol{\theta}}_B = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\text{argzero}}\ \hat{\boldsymbol{\pi}}\left(\boldsymbol{\theta}_0, n\right) - \bar{\boldsymbol{\pi}}\left(\boldsymbol{\theta}, n\right). \tag{12}$$

Interestingly, Theorem 4 implies that $\tilde{\boldsymbol{\theta}}_B$ and $\hat{\boldsymbol{\theta}}$ are equivalent without relying on Assumptions 1, 2 and 4. Therefore, these two methods have the same rates of bias and variance, which are provided in Theorem 4 (under

some conditions). In addition, this allows us to conclude that $\tilde{\boldsymbol{\theta}}_B$ has the same asymptotic distribution as $\hat{\boldsymbol{\theta}}$ (Gouriéroux et al., 1993). Thus, the iterative bootstrap can be thought of as an algorithm to solve the optimization problem of the indirect inference based on H samples of size n.

4 A Simple Example

We illustrate the performance of the various bias correction procedures discussed in the previous sections in a simple simulation example. The objective is to provide empirical illustration of the rates at which the biases of the different methods converge to zero.

We let $X_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ and consider the case where we wish to estimate σ^2 based on the sample $(x_i)_{i=1,\cdots,n}$ via the auxiliary estimator

$$\hat{\pi} = \frac{1}{\log(n)} + \frac{n}{n-1}\hat{\mu}_2 + \frac{10}{n^3}\hat{\mu}_4 + \frac{10}{n^4}\hat{\mu}_6,$$

where $\hat{\mu}_k = n^{-1}\sum_{i=1}^{n} x_i^k$. It is easy to verify that $\hat{\pi}$ is consistent for σ^2 but has a bias given by

$$\mathbb{E}[\hat{\pi}] - \sigma^2 = \frac{1}{\log(n)} + \frac{\sigma^2}{n-1} + \frac{30\sigma^4}{n^3} + \frac{150\sigma^6}{n^4}.$$

We chose this auxiliary estimator to introduce additional bias. Let $\hat{\sigma}^2$, $\tilde{\sigma}^2$, $\hat{\sigma}^2_B$ and $\tilde{\sigma}^2_B$ denote the estimator based on Strategies 1, 2, 3 and 4 respectively. In order to study the rates of convergence of the biases, we consider 10 sample

sizes $n = 5, 10, 15, \ldots, 50$. The absolute values of the biases for the different strategies and sample sizes obtained from 5,000 Monte Carlo replications are presented in Figure 1. It can be observed that Strategies 1 and 4 produce approximately unbiased estimates for sample size larger or equal to 15 while a small bias can be observed for $n = 5$ and 10. Strategy 3 leads to estimates that exhibit a larger bias than the two previously mentioned methods. However, this method corrects most of the bias of $\hat{\pi}$ and appears nearly unbiased when $n > 25$. On the other hand, Strategy 2 corrects only marginally the bias of the auxiliary estimator. It can be observed that the bias of this method drops with n at a rate comparable to the original estimator $\hat{\pi}$. To illustrate further the rates of convergence of the biases against the results of Theorems 1 to 4, we super-imposed curves of the type

$$|\text{bias}_n| = an^{-3} + bn^{-4} + cn^{-5}$$

on the observed biases from Strategies 1 and 4, and curves of the type

$$|\text{bias}_n| = an^{-1} + bn^{-2} + cn^{-3}$$

for Strategies 2 and 3. Specifically, the four estimated curves in Figure 1 are respectively $39.6n^{-3} + 128.7n^{-4} - 938.2n^{-5}$, $9.6n^{-1} - 75.5n^{-2} + 259.2n^{-3}$, $-0.7n^{-1} + 28.2n^{-2} - 6.9n^{-3}$ and $42.1n^{-3} + 67.5n^{-4} + 578.6n^{-5}$ for Strategies 1 to 4. The close fit of these curves with the observed biases clearly illustrates the bias orders provided in the theorems.

5 Application to Regression with Lagged Variables

In the case where the lagged dependent variables are included in a regression model, it is well-known that the Ordinary Least Squares (OLS) estimates are often biased in small samples. Moreover, this bias tends to increase considerably when additional variables are included in the model. There is a vast amount of literature on this subject and a recent discussion can be found in Tanizaki et al. (2006) and the references therein. To illustrate the different bias correction strategies discussed in Section 2 we consider the

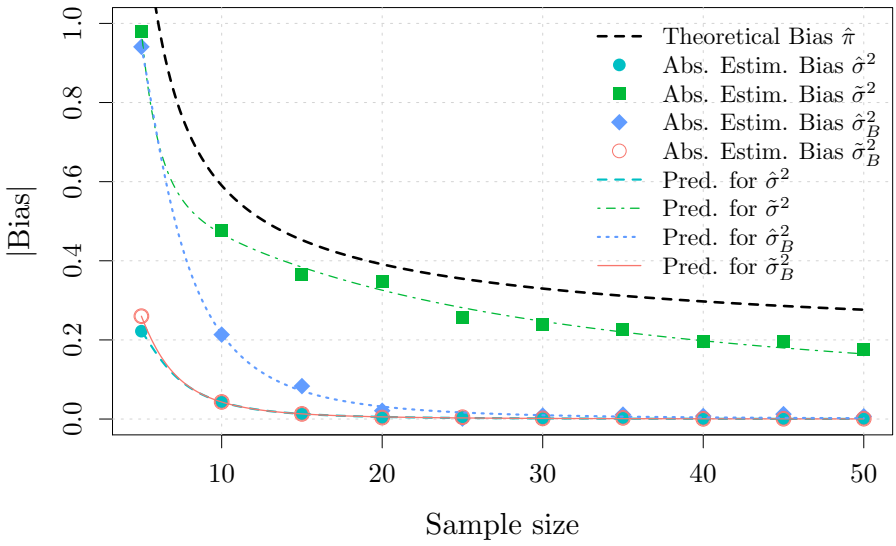

FIGURE 1: *Observed biases from Strategies 1-4 overlaid with curves with leading order n^{-3} for Strategies 1 ($\hat{\sigma}^2$) and 4 ($\tilde{\sigma}^2_B$), and leading order n^{-1} for Strategies 2 ($\tilde{\sigma}^2$) and 3 ($\hat{\sigma}^2_B$), at different n.*

following model:

$$Y_t = \alpha + \beta t + \phi Y_{t-1} + \varepsilon_t,$$

where $t = 1, ..., n$ and $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. We use the following simulation setting: $n = 40$, $\alpha = \beta = 0$, $\phi = 0.9$, $\sigma^2 = 2$ and $H = 100$. The estimated biases and Mean Squared Errors (MSE) of the different approaches for the parameter ϕ are presented in Table 1. Strategies 1 and 4 provide nearly equivalent results and lead to the best performance in terms of both bias and MSE. Indeed, the resulting MSE for these strategies is less than half of the one of the OLS $\hat{\pi}$. In contrast, Strategy 2 does not lead to any improvement compared to the auxiliary estimator (similar bias and MSE). This is expected for consistent auxiliary estimators which is the case for the OLS in this example.

Indeed, for consistent and for sufficiently large H and fixed n, the auxiliary estimator of Strategy 2 allows the following approximation

$$\begin{aligned}\tilde{\boldsymbol{\pi}}(\boldsymbol{\theta}, nH) &= \boldsymbol{\pi}(\boldsymbol{\theta}, nH) + \mathbf{v}(\boldsymbol{\theta}, nH) = \boldsymbol{\theta} + \mathbf{d}(\boldsymbol{\theta}, nH) + \mathbf{v}(\boldsymbol{\theta}, nH) \\ &= \boldsymbol{\theta} + \mathcal{O}_p\left((nH)^{-\min(1,\alpha,\beta)}\right) \approx \boldsymbol{\theta},\end{aligned}$$

where $\alpha > 0$ by Assumption 2 and $\beta > 0$ by Assumption 3 since the auxiliary estimator is consistent. This therefore implies that this method will only provide minor improvement over the original consistent estimator. Strategy 3 is able to correct most of the bias of $\hat{\boldsymbol{\pi}}$ and hence has a comparable MSE to Strategies 1 and 4. Strategies 1 and 4, the overall best performing strategies can be further compared. First, since both resulting estimators

are equivalent (see Theorem 4), their observed sample performance slightly differ for numerical approximation's reasons. Second, the calculation speed is expected to be significantly slower for Strategy 1 which involves finding the root of an equation which in turn involves the computation of H times the auxiliary estimator, while Strategy 4 only needs one computation of the auxiliary estimator at each step.

	Parameter ϕ			
	Bias		MSE	
$\hat{\pi}$ (OLS)	$-1.51 \cdot 10^{-1}$	$(4.73 \cdot 10^{-3})$	$3.45 \cdot 10^{-2}$	$(2.05 \cdot 10^{-3})$
Strategy 1	$-1.93 \cdot 10^{-2}$	$(4.97 \cdot 10^{-3})$	$1.40 \cdot 10^{-2}$	$(1.25 \cdot 10^{-3})$
Strategy 2	$-1.52 \cdot 10^{-1}$	$(4.81 \cdot 10^{-3})$	$3.48 \cdot 10^{-2}$	$(2.06 \cdot 10^{-3})$
Strategy 3	$-5.52 \cdot 10^{-2}$	$(5.30 \cdot 10^{-3})$	$1.68 \cdot 10^{-2}$	$(1.38 \cdot 10^{-3})$
Strategy 4	$-2.98 \cdot 10^{-2}$	$(5.62 \cdot 10^{-3})$	$1.46 \cdot 10^{-2}$	$(1.44 \cdot 10^{-3})$

TABLE 1: *Estimated biases and MSE for Strategies 1 to 4 based on 500 Monte Carlo replications for the setting described in Section 5. The numbers in parentheses are the standard deviations obtained by nonparametric bootstrap.*

6 Application to Robust Estimation of Income Distribution

We now consider robust estimation of parameters in income distribution models, which are often used to estimate income inequality (Cowell, 2011) and to compare different distributions. The benefits of robust methods in this context were already demonstrated in the literature, see e.g. Cowell and Victoria-Feser (2007) and the references therein.

Let $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ denote the parameter vector of an income distribution $F_{\boldsymbol{\theta}_0}$ that we wish to estimate. In the robustness paradigm, one is concerned with the possibility that the data generating distribution is not exactly the postulated model $F_{\boldsymbol{\theta}_0}$ but rather a deviation model of the form

$$(1-\varepsilon)F_{\boldsymbol{\theta}_0} + \varepsilon F_C, \tag{13}$$

where F_C is an unspecified contamination distribution function and $\varepsilon \in [0,1]$ is the amount of deviation from the postulated model $F_{\boldsymbol{\theta}_0}$. To estimate the income distribution $F_{\boldsymbol{\theta}_0}$ and use it to study aspects of income inequality such as poverty, etc., a key component is to understand how $\varepsilon > 0$ affects the estimation of $\boldsymbol{\theta}_0$ and hence the conclusions of the study. If ε is large, which means that a large proportion of the data are generated from F_C, an arbitrary contamination distribution other than $F_{\boldsymbol{\theta}_0}$, then we may conclude that the analysis of the model $F_{\boldsymbol{\theta}_0}$ will not yield useful results. However, if ε is relatively small, we may expect the analysis of $F_{\boldsymbol{\theta}_0}$ to be indicative. In other words, the analysis should have certain robustness under slight model deviation.

A thorough review of possible parametric models $F_{\boldsymbol{\theta}_0}$ for income distributions can be found in Kleiber and Kotz (2003). Robust estimators for income distributions were first proposed by Victoria-Feser and Ronchetti (1994) and Victoria-Feser and Ronchetti (1997) for grouped data. These robust estimators are in fact Optimal B-Robust Estimators (OBRE) of Hampel et al.

(1986) and are the most efficient estimators among the class of M-estimators with bounded Influence Function (IF).

The OBRE estimator is the root of the estimating equation

$$\sum_{i=1}^{n} \mathbf{A}(\boldsymbol{\theta}) \left[\boldsymbol{s}(x_i;\boldsymbol{\theta}) - a(\boldsymbol{\theta})\right] w(x_i;\boldsymbol{\theta},c) = 0, \tag{14}$$

where $\mathbf{A}(\boldsymbol{\theta})$ is a $p \times p$ matrix, often chosen to maximize the estimation efficiency, $a(\boldsymbol{\theta})$ is a consistency correction that satisfies

$$\int \mathbf{A}(\boldsymbol{\theta}) \left[\boldsymbol{s}(x;\boldsymbol{\theta}) - a(\boldsymbol{\theta})\right] w(x;\boldsymbol{\theta},c) dF_{\boldsymbol{\theta}}(x) = 0, \tag{15}$$

where $\boldsymbol{s}(x;\boldsymbol{\theta}) \equiv \partial \log\{f(x;\boldsymbol{\theta})\}/\partial\boldsymbol{\theta}$ is the score function with $f(x;\boldsymbol{\theta})$ the probability density function of $F_{\boldsymbol{\theta}}$, and $w(x;\boldsymbol{\theta},c)$ is a weight function that bounds $\mathbf{A}(\boldsymbol{\theta})\left[\boldsymbol{s}(x;\boldsymbol{\theta}) - a(\boldsymbol{\theta})\right]$, such as the Huber or the Bi-square weight function (see e.g. Hampel et al., 1986). If $\mathbf{A}(\boldsymbol{\theta})$ is set to identity and the consistency constraint is integrated in the estimating equation, the OBRE estimator is the root of the estimating equation

$$\sum_{i=1}^{n} \boldsymbol{s}(x_i;\boldsymbol{\theta}) w(x_i;\boldsymbol{\theta},c) - \frac{\sum_{i=1}^{n} w(x_i;\boldsymbol{\theta},c)}{\int w(x;\boldsymbol{\theta},c) dF_{\boldsymbol{\theta}}(x)} \int \boldsymbol{s}(x;\boldsymbol{\theta}) w(x;\boldsymbol{\theta},c) dF_{\boldsymbol{\theta}}(x) = 0. \tag{16}$$

The OBRE estimator obtained from (16) has the property that it is consistent if $\varepsilon = 0$, and is approximately consistent even if $\varepsilon > 0$, hence is typically favored in the robust statistics literature. However, the OBRE is in general difficult to compute, mainly because the integrals in (16) are not analytically

tractable for income distribution models. Due to these numerical difficulties, the OBRE is hardly applicable for "complex" densities.

As in Guerrier et al. (2017), we propose to use a simpler robust starting estimator that is inconsistent but easy to compute and to correct for consistency by means of indirect inference. This approach has actually been used in Moustaki and Victoria-Feser (2006) for robust estimation of GLLVM. The starting estimator we propose to use is a weighted MLE $\hat{\boldsymbol{\pi}}$, which is the solution of

$$\sum_{i=1}^{n} w(x_i; \boldsymbol{\pi}, c)\boldsymbol{s}(x_i; \boldsymbol{\pi}) = \mathbf{0}, \tag{17}$$

where $w(x_i; \boldsymbol{\pi}, c)$ are weights such that $w(x_i; \boldsymbol{\pi}, c)\boldsymbol{s}(x_i; \boldsymbol{\pi})$ is bounded for all i. This guarantees a bounded IF for the subsequent consistent estimators obtained from indirect inference or iterative bootstrap as shown in Genton and Ronchetti (2003).

To illustrate the finite sample properties of the different bias correction strategies for parameter estimation in an income distribution using the weighted MLE (17) as starting estimator, we consider the case of the Lomax distribution with density

$$f(x) = \frac{q}{b}\left(1 + \frac{x}{b}\right)^{-(q+1)}, \quad x > 0.$$

To demonstrate the performance of the various estimators, we simulated data from three different settings. In the first setting, we used a sample size of $n = 50$, which is relatively small, while in the second setting, the sample size

is $n = 1000$. In the third setting, we used the same sample size $n = 1000$, but contaminated a randomly chosen 1% of the data which are multiplied by 1000. In all settings, the true parameters are $b = 5$ and $q = 2$. We compute the weighted MLE $\hat{\boldsymbol{\pi}}$ given in (17) which is inconsistent, then implemented the two versions of the indirect inference estimators described in Strategies 1 and 2, using $\hat{\boldsymbol{\pi}}$ as the initial estimator. These estimators are denoted $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$ respectively. We also implemented the MLE $\hat{\boldsymbol{\pi}}_{ML}$ as a benchmark estimator. Here, we set $H = 100$ throughout.

In the first row of Figure 2, when the sample size is 50, the weighted MLE has a large bias that cannot be completely eliminated via the one sample indirect inference estimator. This bias can only be corrected by the indirect inference estimator with H samples of size n. In this case, even the MLE has a clear bias, since the asymptotic consistency of MLE does not demonstrate at such small sample. This is a well known issue, and in the case of the Lomax distribution, Giles et al. (2013) derive the analytical expression of the finite sample bias. However, analytical correction for finite sample bias of the MLE is by far not available for all income distribution whereas simulation based bias correction methods can always be applied. In the second row of Figure 2, with a large sample size ($n = 1000$), the sample bias of the weighted MLE is corrected by both Strategies 1 and 2. In this case, the performance of both indirect inference estimators is similar to the asymptotically optimal estimator MLE. Finally, when data is contaminated as in the third row of Figure 2, both indirect estimators can correct the

bias and result in robustness. Not surprisingly, the benchmark estimator MLE in this case has clear bias, since MLE is not a robust estimator. We emphasized that the robust estimation approach taken here is not specific to income distributions. It can be used for any model for which a score function can be specified. This opens the door to simple computation of robust estimators especially for complex models.

Finally, we reconsider the scenarios presented in Figure 2 with the intention of evaluating the accuracy of the estimated variance $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$. Indeed, a "natural" estimator of these quantity can be obtained using the asymptotic normality results of Gouriéroux et al. (1993) (see also Genton and Ronchetti, 2003). Under typical regularity assumptions and assuming $H \to \infty$, this result allows the construction of a consistent estimator for the covariance matrix of both $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$ as:

$$\widehat{\text{var}}(\hat{\boldsymbol{\theta}}) = \widehat{\text{var}}(\tilde{\boldsymbol{\theta}}) = \mathbf{K}^{-1}\widehat{\text{var}}(\hat{\boldsymbol{\pi}})\mathbf{K}^{-T}, \tag{18}$$

where $\mathbf{K}$ denotes the Jacobian matrix of the transformation estimated by numerical derivative and $\widehat{\text{var}}(\hat{\boldsymbol{\pi}})$ denotes the covariance matrix of $\hat{\boldsymbol{\pi}}$ obtained by parametric bootstrap, quantity which is in particular directly available from the output of Strategy 1.

Table 2 presents the empirical coverage rate of the estimated parameters based on 1000 Monte Carlo replications for the three scenarios presented in Figure 2. For the MLE, we use the plug-in estimator of the asymptotic ex-

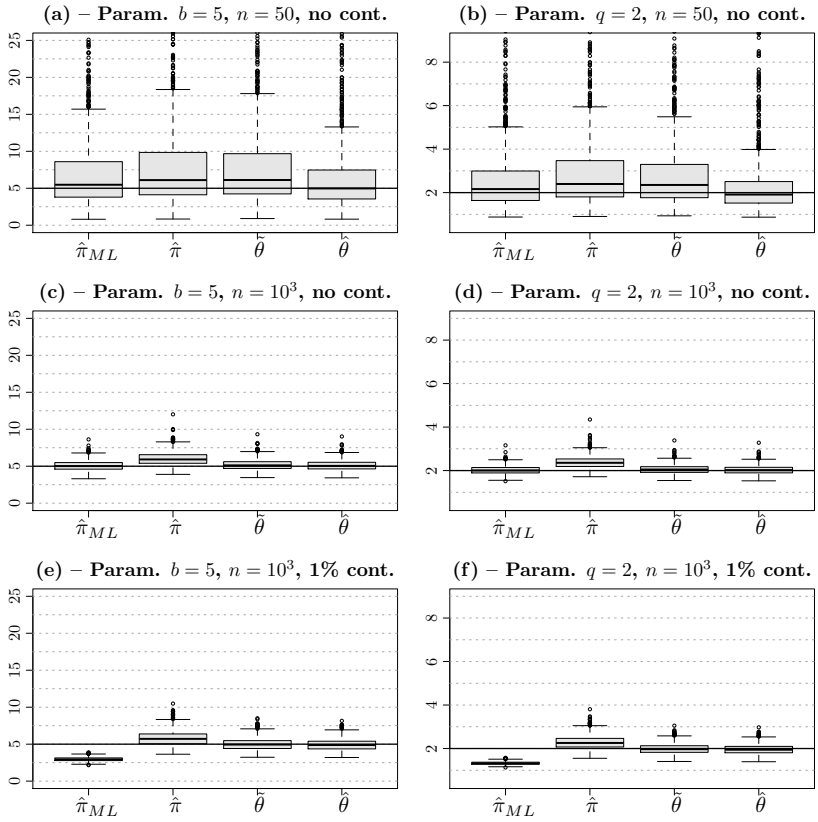

FIGURE 2: *Finite sample distribution of the estimators of b (left panel) and q (right panel) for the Lomax simulation study. The true parameters are* $b = 5$ *and* $q = 2$, $\hat{\boldsymbol{\pi}}_{ML}$ *is the MLE,* $\hat{\boldsymbol{\pi}}$ *is the weighted MLE given in (17),* $\hat{\boldsymbol{\theta}}$ *and* $\tilde{\boldsymbol{\theta}}$ *are the indirect estimators using respectively Strategies 1 and 2 based on* $\hat{\boldsymbol{\pi}}$ *with* $H = 100$. *The first row corresponds to samples of size* $n = 50$, *the second row to samples of size* $n = 1000$, *and the third row to samples of size* $n = 1000$, *in which 1% of the data (chosen randomly) are multiplied by* 1000.

pression. As expected, the results of Strategies 1 and 2 are very close since the estimators have similar performance and their confidence intervals are both based on (18). In the small sample case (i.e. $n = 50$), the estimators based on indirect inference provide far better coverage probability when compared to the MLE. This can be explained partially by the fact that (18) uses

a parametric bootstrap estimator $\widehat{\text{var}}(\hat{\boldsymbol{\pi}})$ while the covariance matrix of MLE is based on the asymptotic expression. In the large sample case ($n = 1000$), all methods have similar performance while MLE appears slightly better. This is probably due to the variability introduced in (18) by the numerical approximation of $\mathbf{K}$. In the contaminated scenario, the performance of the indirect inference-based methods is slightly hampered due to the extreme outliers, and as expected, the MLE completely fails.

	Coverage 90%		Coverage 95%		Coverage 99%	
	b	q	b	q	b	q
$n = 50$						
Strategy 1	88.8	90.2	91.2	91.5	94.5	95.6
Strategy 2	86.5	88.2	92.4	91.2	94.6	94.3
MLE	21.9	21.9	25.1	25.9	32.9	32.9
$n = 1000$						
Strategy 1	92.4	90.3	96.1	96.2	99.4	99.1
Strategy 2	91.2	91.9	96.4	95.3	99.3	99.3
MLE	89.5	89.5	95.1	95.3	99.2	98.8
$n = 1000$ (C)						
Strategy 1	88.2	88.8	92.5	91.6	97.2	98.3
Strategy 2	89.5	89.0	92.6	92.1	97.5	97.8
MLE	0.0	0.0	0.0	0.0	0.1	0.0

TABLE 2: *Estimated coverage probabilities for the parameters b and q in the Lomax distribution. The scenario $n = 1000$ (C) denotes the case where we consider a sample size of $n = 1000$, in which 1% of the data (chosen randomly) are multiplied by 1000. Results based on 1000 Monte Carlo replications ($H = 100$).*

7 Application to Generalized Linear Latent Variable Models

GLLVM (Bartholomew, 1984 and Moustaki and Knott, 2000) are very popular in various areas of research such as psychology, social sciences or economics. They link theoretical concepts that are not observable to manifest variables that are observed. When the manifest variables fall outside the normal model, analysis of such model deviates from factor analysis, and the estimation method based on the maximum likelihood becomes computationally very challenging. This is because not only the number of parameters to be estimated increases very quickly with the increasing complexity of the model, but also the models include non-observable latent variables that need to be marginalized out from the likelihood function, which results in complicated multidimensional integrals. The numerical approximation of these integrals includes adaptive quadratures (see e.g. Rabe-Hesketh et al., 2002) or the Laplace approximation (see e.g. Huber et al., 2004).

Even with these approximations, maximizing the approximated likelihood function remains numerically quite challenging. An alternative is to replace the likelihood with composite likelihood (Lindsay, 1988). Composite likelihood functions are formed by the product of individual component likelihoods, each corresponding to marginal or conditional events. The pairwise composite likelihood function is an example of a composite likelihood function. If the elements are properly chosen, the maximum composite likelihood

estimator is consistent and normally distributed. Although composite likelihood is constructed as a compromise when likelihood is too hard to form or to compute, the computational challenge of composite likelihood estimator can nevertheless be still substantial for use as an auxiliary estimator for GLLVM, see for example Varin (2008), Varin et al. (2011) and Katsikatsou et al. (2012) for GLLVM with ordinal variables.

Another instance of the latent variable models with challenging computation is when the variances have a dynamic autoregressive component. These models have been used widely. For example, Diebold and Nerlove (1989) used it in multivariate autoregressive conditional heteroskedasticity (ARCH) time-series. Sentana et al. (2008) proposed to use indirect inference to allow the estimation of any state space model with generalized autoregressive conditional heteroskedasticity (GARCH) disturbances, using the model proposed by Harvey et al. (1992) as auxiliary model.

In this application, we build a simulation based estimator for the parameters of GLLVM with ordinal manifest variables. The first step of our estimator is defined by an algorithm that provides an easy to compute but inconsistent estimator, and iterative bootstrap is then used to remove its bias. The model we consider in this section includes 90 parameters in total to illustrate the applicability of the bias correction techniques presented in Section 2 to complex settings. For comparison, we also used Mplus (see e.g. Muthén and Muthén, 1998-2011), a standard software for GLLVM, where the likelihood function is approximated using adaptive Gauss quadratures

to approximate integrals. However, for other latent variable models where software such as Mplus is not available, our procedure can still be readily implemented, as long as one can construct a starting estimator, which does not need to have an explicit form and does not need to be consistent.

Let $z^{(k)}$, $k = 1, \ldots, q$ be the latent variables and $x^{(j)}$, $j = 1, \ldots, p$, be the manifest variables, $p > q$. A crucial assumption is that all the dependence structure of the manifest variables is explained by the latent variables. Thus, conditionally on the latent variables, the manifest variables are independent of each other with conditional density $g_j(x^{(j)}|\mathbf{z})$. Hence the joint conditional density of the manifest variables is $\prod_{j=1}^{p} g_j(x^{(j)}|\mathbf{z})$. In GLLVM, g_j is assumed to belong to the exponential family but can be different for different j, and it satisfies

$$\nu_j(\mathbb{E}(x^{(j)}|\mathbf{z}_{(2)})) = \boldsymbol{\lambda}^{(j)\mathrm{T}}\mathbf{z}, \tag{19}$$

where

$$\mathbf{z} = (1, \mathbf{z}_{(2)}^{\mathrm{T}})^{\mathrm{T}},\ \mathbf{z}_{(2)} = (z^{(1)}, \ldots, z^{(q)})^{\mathrm{T}},\ \boldsymbol{\lambda}^{(j)} = (\lambda_0^{(j)}, \ldots, \lambda_q^{(j)})^{\mathrm{T}} = (\lambda_0^{(j)}, \boldsymbol{\lambda}_{(2)}^{(j)\mathrm{T}})^{\mathrm{T}},$$

and where $\boldsymbol{\lambda}_{(2)}^{(j)}$ are the loadings, and ν_j are the link functions determined by the corresponding g_j, $j = 1, \ldots, p$. The density of the latent variables, denoted by $h(\mathbf{z}_{(2)})$ is assumed to be multivariate standard normal. This implies that the latent variables are independent of each other. The joint distribution of the manifest and latent variables is then $\prod_{j=1}^{p} g_j(x^{(j)}|\mathbf{z})h(\mathbf{z}_{(2)})$,

and the marginal density for the manifest variables is

$$f_{\boldsymbol{\lambda},\boldsymbol{\phi}}(\mathbf{x}) = \int \dots \int \Big\{ \prod_{j=1}^{p} g_j(x^{(j)}|\mathbf{z}) \Big\} h(\mathbf{z}_{(2)}) d\mathbf{z}_{(2)}. \tag{20}$$

Given a sample of n observations $\boldsymbol{x}_1, \dots, \boldsymbol{x}_\mathrm{n}$ where $\boldsymbol{x}_\mathrm{i} = (x_i^{(1)}, \dots, x_i^{(p)})^T$, $i = 1, \dots, n$, the log-likelihood of the loadings $\boldsymbol{\lambda}$ and the scale parameters $\boldsymbol{\phi}$ is

$$\begin{aligned} l(\boldsymbol{\lambda}, \boldsymbol{\phi}|\mathbf{x}) &= \sum_{i=1}^{n} \log f_{\boldsymbol{\lambda},\boldsymbol{\phi}}(\mathbf{x}_i), \\ &= \sum_{i=1}^{n} \log \int \cdots \int \prod_{j=1}^{p} \exp\Big\{ \frac{x_i^{(j)} u_j(\boldsymbol{\lambda}^{(j)\mathrm{T}}\mathbf{z}) - b_j(u_j(\boldsymbol{\lambda}^{(j)\mathrm{T}}\mathbf{z}))}{\phi_j} \\ &\quad + c_j(x^{(j)}, \phi_j) \Big\} h(\mathbf{z}_{(2)}) d\mathbf{z}_{(2)}. \end{aligned} \tag{21}$$

The MLE for $\boldsymbol{\lambda}$ and $\boldsymbol{\phi}$ is then obtained by maximizing (21) with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\phi}$, using approximations to the integrals such as adaptive quadratures or Laplace approximation.

Huber et al. (2004) applied the Laplace method to GLLVM and obtained the following approximate log-likelihood function

$$\tilde{l}(\boldsymbol{\lambda}, \boldsymbol{\phi}|\mathbf{x}_i) = \sum_{i=1}^{n} \big\{ -\log\left(\det\left[\boldsymbol{\Gamma}(\boldsymbol{\lambda}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i)\right]\right)/2 + p\, Q(\boldsymbol{\lambda}, \boldsymbol{\phi}, \hat{\mathbf{z}}_i, \mathbf{x}_i) \big\}, \tag{22}$$

where

$$\boldsymbol{\Gamma}(\boldsymbol{\lambda}, \boldsymbol{\phi}, \mathbf{z}) = -p \frac{\partial^2 Q(\boldsymbol{\lambda}, \boldsymbol{\phi}, \mathbf{z}, \mathbf{x})}{\partial \mathbf{z}^{\mathrm{T}} \partial \mathbf{z}},$$

$$Q(\boldsymbol{\lambda}, \boldsymbol{\phi}, \mathbf{z}, \mathbf{x}) = \frac{1}{p} \sum_{j=1}^{p} \left(\frac{x^{(j)} u_j(\boldsymbol{\lambda}^{(j)\mathrm{T}} \mathbf{z}) - b(u_j(\boldsymbol{\lambda}^{(j)\mathrm{T}} \mathbf{z}))}{\phi_j} + c_j(x^{(j)}, \phi_j) \right) - \frac{\mathbf{z}_{(2)}^{\mathrm{T}} \mathbf{z}_{(2)}}{2p} - \frac{q}{2p} \log(2\pi), \tag{23}$$

and $\hat{\mathbf{z}}$ is the maximum of $Q(\boldsymbol{\lambda}, \boldsymbol{\phi}, \mathbf{z}, \mathbf{x})$. Ignoring the first summand in the approximate log-likelihood in (22), we obtain

$$l^{\star}(\boldsymbol{\lambda}, \boldsymbol{\phi}, \mathbf{z}_i | \mathbf{x}_i) = p \sum_{i=1}^{n} Q(\boldsymbol{\lambda}, \boldsymbol{\phi}, \mathbf{z}_i, \mathbf{x}_i), \tag{24}$$

which corresponds to the h-likelihood proposed by Lee and Nelder (1996) for GLMM in the context of latent variable models. Maximizing (24) yields the penalized quasi-likelihood (PQL) estimator proposed in the context of GLMM (Breslow and Clayton, 1993). Indeed, maximizing (24) provides the same estimators that would result from the maximization of a "likelihood" where the latent scores $\mathbf{z}_i$ are considered as parameters.

If the $\mathbf{z}_i$'s were considered as fixed in (24), the maximum of $l^{\star}$ over $\boldsymbol{\lambda}$ and $\boldsymbol{\phi}$ would be the same as the MLE of a Generalized Linear Model (GLM), in which the covariates would be the $\mathbf{z}_i$'s. Therefore, as a starting estimator, we propose to find estimates for the loadings and the scale parameter by fitting a GLM on the data with predictions of the latent scores as covariates. These predictions for the $\mathbf{z}_i$ are found in a premilinary step where we perform a Factor Analysis on the data as if they were normal and compute the Bartlett predictions of the latent scores (Bartlett, 1950). A similar idea has been used

in Sardy and Victoria-Feser (2012). This initial estimator is thus obviously not consistent but is easy to compute.

When the manifest variables are ordinal, their conditional distribution $g_j(x^{(j)}|\mathbf{z})$ is the multinomial distribution. Specifically, for j, let $m^{(j)}$ be the number of categories and $\gamma_s^{(j)}$ be the probability that the manifest variable $x^{(j)}$ is smaller or equal to s, where $s = 1, \ldots, m^{(j)} - 1$. We set a proportional odds model

$$\gamma_s^{(j)} = \frac{e^{\boldsymbol{\lambda}_s^{(j)\mathrm{T}}\mathbf{z}}}{1 + e^{\boldsymbol{\lambda}_s^{(j)\mathrm{T}}\mathbf{z}}}, \tag{25}$$

where $\boldsymbol{\lambda}_s^{(j)} = (\lambda_{0s}^{(j)}, \boldsymbol{\lambda}_{(2)}^{(j)\mathrm{T}})^{\mathrm{T}} = (\lambda_{0s}^{(j)}, \lambda_1^{(j)}, \ldots, \lambda_q^{(j)})^{\mathrm{T}}$, $\lambda_{0s}^{(j)}$ being the threshold for the sth category with

$$-\infty < \lambda_{01}^{(j)} \leq \lambda_{02}^{(j)} \leq \ldots \lambda_{0(m^{(j)}-1)}^{(j)} < \infty.$$

In the simulation setting presented here we consider three latent variables ($q = 3$) and fifteen ordinal manifest variables ($p = 15$), each with five categories ($m_j = 5$). The true parameter values are given in Appendix F (in the Supplementary Materials). The 15 nil loadings are fixed and the thresholds are not constrained to have the same values across manifest variables. This yields $(m_j - 1 + q) \times p - 15 = 90$ parameters to estimate. For this model, the starting inconsistent estimator $\hat{\boldsymbol{\pi}}$ is obtained by estimating an ordinal GLM in which the covariates are the Bartlett scores obtained as explained before. The computational details are given in Appendix F (in the Supplementary Materials). We generate 120 samples of size $n = 300$ and use $H = 100$ for

the iterative bootstrap procedure.

We compare our estimator ($\tilde{\boldsymbol{\theta}}_B$) in terms of bias distribution to the MLE obtained via Mplus ($\hat{\boldsymbol{\theta}}_{ML}$) which uses adaptive quadratures to approximate the integrals in (21). Figure 3 shows the bias distribution of the estimates of four loadings (the $\lambda_k^{(j)}$'s). For other loading estimates, the results are very similar. As expected, the starting estimator ($\hat{\boldsymbol{\pi}}$) is biased, but applying the iterative bootstrap provides a nearly unbiased estimator ($\tilde{\boldsymbol{\theta}}_B$), which is comparable to the one obtained using adaptive quadratures to approximate the likelihood function ($\hat{\boldsymbol{\theta}}_{ML}$), with a slight loss of efficiency. The same can be said when we inspect the threshold estimates (the $\lambda_{0s}^{(j)}$'s), see Figure 4. As a last point, we would like to comment on the computational advantage of the iterative bootstrap method. Indeed, we have also attempted to correct the bias of the initial estimator with indirect inference in Strategies 1 and 2, but the optimizations encountered difficulties and did not converge. Since the iterative bootstrap can be seen as an algorithm to calculate the indirect inference estimator in Strategy 1 (see Theorem 4), the former can then be favored over the direct optimization of the indirect inference estimator definition, from a computational point of view, especially in complex model settings such as GLLVM.

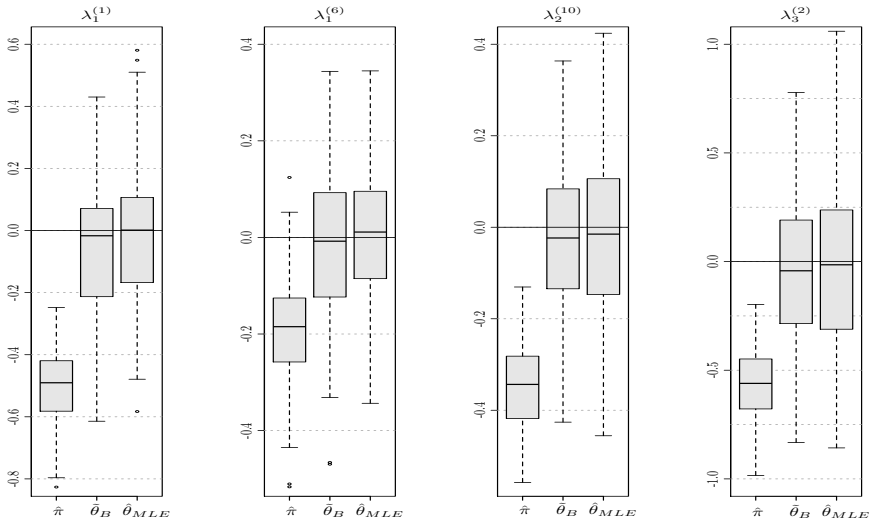

FIGURE 3: *Empirical distribution of the estimates, centered at the true value, of four loadings obtained by the starting inconsistent estimator ($\hat{\boldsymbol{\pi}}$), the estimator based on iterative bootstrap ($\tilde{\boldsymbol{\theta}}_B$) and the approximated MLE obtained using Mplus ($\hat{\boldsymbol{\theta}}_{ML}$).*

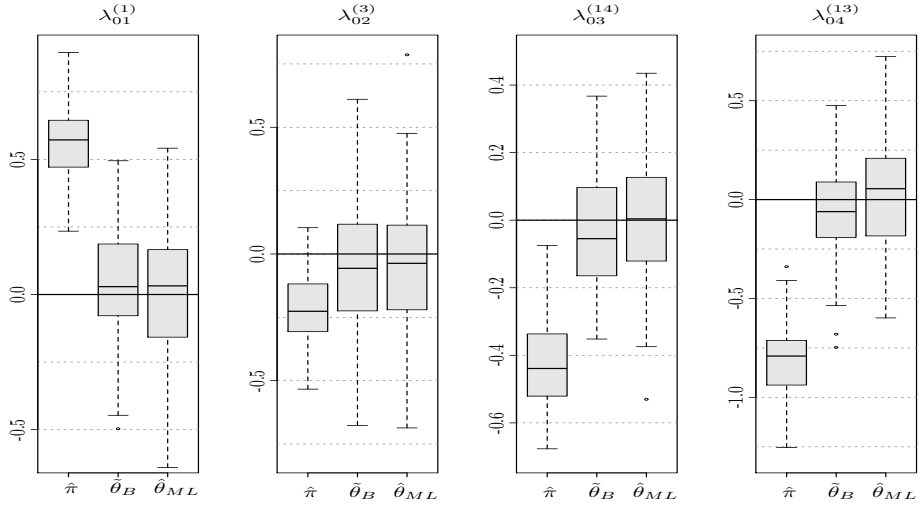

FIGURE 4: *Empirical distribution of the bias of the estimates, centered at the true value, of four thresholds obtained by the starting inconsistent estimator ($\hat{\boldsymbol{\pi}}$), the estimator based on iterative bootstrap ($\tilde{\boldsymbol{\theta}}_B$) and the MLE obtained using Mplus ($\hat{\boldsymbol{\theta}}_{ML}$).*

8 Conclusions

In this paper, we extend and provide refinements on the known properties of four bias correction strategies. These results allow to build a framework defining simulation-based estimators that is particularly relevant when considering complex models. Indeed, based on a simple and/or computationally efficient estimator that is typically biased and possibly inconsistent, several simulation based methods can be used to define new estimators that are both consistent and benefit from a reduced finite sample bias.

In particular, we discover and demonstrate the equivalence of one version of indirect inference (i.e. Strategy 1) and the iterative bootstrap, which both typically correct finite sample biases up to the order n^{-3}. This shows that iterative bootstrap can be considered as a computationally efficient algorithm to implement the indirect inference. Moreover, it deepens our understanding of the properties of the iterative bootstrap, which is consequently consistent and asymptotically normally distributed under the additional assumptions used in the indirect inference framework (see Gouriéroux et al., 1993 for details). The simulation studies presented in Sections 4 and 5 empirically confirm this finding and show that the two methods have nearly identical finite sample performance.

On the computational aspects, as we discussed in Section 7, the advantage of the iterative bootstrap procedure is that it is faster to compute and can be used in complex models such as the GLLVM where the indirect inference

	Ensures consistency	Finite sample bias after correction	Computational Burden	Numerical issues (in complex models)
Strategy 1	Yes	Very small	High	Likely
Strategy 2	Yes	Similar to $\hat{\pi}$	High	Likely
Strategy 3	No	Small	Low	Unlikely
Strategy 4	Yes	Very small	Medium	Unlikely

TABLE 3: *Comparative table of the merits and disadvantages of the four strategies discussed in Section 2. Strategies 1 and 2 correspond to two versions of indirect inference, Strategy 3 denotes the (one step) bootstrap bias corrected estimator and Strategy 4 corresponds to the iterative bootstrap method.*

method even fails to numerically converge.

Finally, between the two indirect inference procedures, Strategy 1 is superior to Strategy 2 when finite sample bias is an issue and sample size is small, or when sample size is large but it is still faster to compute $\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}, n)$ H times than to compute $\hat{\boldsymbol{\pi}}(\boldsymbol{\theta}, nH)$ once. Otherwise, Strategy 2 will be recommended. We provide a summary of the merits and disadvantages of the different strategies in Table 3. Overall, the iterative bootstrap is often the best choice.

References

Arvanitis, S. and Demos, A. On the Validity of Edgeworth Expansions and Moment Approximations for Three Indirect Inference Estimators. *Working paper No. 1406, Dept. of IEES, AUEB*, 2013.

Arvanitis, S. and Demos, A. Valid Locally Uniform Edgeworth Expansions for a Class of Weakly Dependent Processes or Sequences of Smooth Transformations. *Journal of Time Series Econometrics*, 6(2):183–235, 2014.

Arvanitis, S. and Demos, A. A Class of Indirect Inference Estimators: Higher-Order Asymptotics and Approximate Bias Correction. *The Econometrics Journal*, 18(2):200–241, 2015.

Bartholomew, D. J. The foundations of factor analysis. *Biometrika*, 71: 221–232, 1984.

Bartlett, M. S. Tests of significance in factor analysis. *British Journal of Statistical Psychology*, 3:77–85, 1950.

Breslow, N. E. and Clayton, D. G. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88:9–25, 1993.

Cowell, F. A. *Measuring Inequality*. Oxford University Press, Oxford, third edition, 2011.

Cowell, F. A. and Victoria-Feser, M.-P. Robust stochastic dominance: A semi-parametric approach. *Journal of Economic Inequality*, 7: 21–37, 2007.

Diebold, F. and Nerlove, M. The dynamics of exchange rate volatility: a multivariate latent factor arch model. *Journal of Applied econometrics*, 4:1–21, 1989.

Dupuis, D. J. and Morgenthaler, S. Robust weighted likelihood estimators with an application to bivariate extreme value problems. *Canadian Journal of Statistics*, 30:17–36, 2002.

Durbin, J. and Koopman, S. J. *Time series analysis by state space methods*. Oxford University Press, 2001.

Efron, B. and Tibshirani, R. *An introduction to the bootstrap*, volume 57. CRC press, 1993.

Gallant, A. R. and Tauchen, G. Which moments to match? *Econometric Theory*, 12:657–681, 1996a.

Gallant, A. R. and Tauchen, G. Which moments to match. *Econometric Theory*, 12:657–681, 1996b.

Genton, M. G. and Ronchetti, E. Robust indirect inference. *Journal of the American Statistical Association*, 98:1–10, 2003.

Giles, D. E. and Feng, H. and Godwin, R. T. On the bias of the maximum likelihood estimator for the two-parameter lomax distribution. *Communications in Statistics-Theory and Methods*, 42:1934–1950, 2013.

Gouriéroux, C. and Monfort, A. *Simulation-based econometric methods*. Oxford University Press, 1997.

Gouriéroux, C. and Monfort, A. and Renault, E. Indirect inference. *Journal of applied econometrics*, 8:S85–S118, 1993.

Gouriéroux, C. and Renault, E. and Touzi, N. Calibration by simulation for small sample bias correction. In R. Mariano, T. Schuermann, and M. J. Weeks, editors, *Simulation-based inference in econometrics: Methods and applications*, pages 328–358. Cambridge University Press, 2000.

Guerrier, S. and Orso, S. and Victoria-Feser, M.-P. Bounded-influence robust estimation of copulas. Working paper, University of Geneva, 2017.

Hampel, F. R. and Ronchetti, E. and Rousseeuw, P. J. and Stahel, W. A. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley, New York, 1986.

Harvey, A. and Ruiz, E. and Sentana, E. Unobserved component time series models with arch disturbances. *Journal of Econometrics*, 52: 129–157, 1992.

Harvey, A. C. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, 1990.

Huber, P. and Ronchetti, E. and Victoria-Feser, M.-P. Estimation of generalized linear latent variable models. *Journal of the Royal Statistical Society Series B*, 66:893–908, 2004.

Katsikatsou, M. and Moustaki, I. and Yang-Wallentin, F. and Jreskog, K. G. Pairwise likelihood estimation for factor analysis models with ordinal data. *Computational Statistics & Data Analysis*, 56:4243–4258, 2012.

Kleiber, C. and Kotz, S. *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley series in probability and statistics. Wiley, New York, 2003.

Kuk, A. Y. C. Asymptotically unbiased estimation in generalized linear models with random effects. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57:395–407, 1995.

Lee, Y. and Nelder, J. A. Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58:619–678, 1996.

Lindsay, B. Composite likelihood methods. *Contemporary Mathematics*, 80:221–239, 1988.

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

F. Mealli and C. Rampichini. Estimating binary multilevel models through indirect inference. *Computational Statistics & Data Analysis*, 29: 313 – 324, 1999.

Moustaki, I. and Knott, M. Generalized latent trait models. *Psychometrika*, 65:391–411, 2000.

Moustaki, I. and Victoria-Feser, M.-P. Bounded-bias robust inference for generalized linear latent variable models. *Journal of the American Statistical Association*, 101:644–653, 2006.

Muthén, L. K. and Muthén, B. O. *Mplus User's Guide. Fifth Edition*, volume 57. Los Angeles, CA: Muthén & Muthén, 1998-2011.

Newey, W. K. and McFadden, D. Large Sample Estimation and Hypothesis Testing. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4, pages 2111–2245. Elsevier, 1994.

Rabe-Hesketh, S. and Skrondal, A. and Pickles, A. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2:1–21, 2002.

Sardy, S. and Victoria-Feser, M.-P. Isotone additive latent variable models. *Statistics and Computing*, 22:647–659, 2012.

Sentana, E. and Calzolari, G. and Fiorentini, G. Indirect estimation of large conditionally heteroskedastic factor models, with an application to the dow 30 stocks. *Journal of Econometrics*, 146:10–25, 2008.

Smith, A. A. Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8:S63–S84, 1993.

Tanizaki, H. and Hamori, S. and Matsubayashi, Y. On least-squares bias in the ar (p) models: bias correction using the bootstrap methods. *Statistical Papers*, 47:109–124, 2006.

Varin, C. On composite marginal likelihoods. *Advances in Statistical Analysis*, 92:1–28, 2008.

Varin, C. and Reid, N. and Firth, D. An overview of composite likelihood methods. *Statistica Sinica*, 21:1–41, 2011.

Victoria-Feser, M.-P. and Ronchetti, E. Robust methods for personal income distribution models. *The Canadian Journal of Statistics*, 22: 247–258, 1994.

Victoria-Feser, M.-P. and Ronchetti, E. Robust estimation for grouped data. *Journal of the American Statistical Association*, 92:333–340, 1997.