REVSTAT – Statistical Journal
Volume 5, Number 1, March 2007, 85–96

DE-BIASING WEIGHTED MLE VIA INDIRECT INFERENCE: THE CASE OF GENERALIZED LINEAR LATENT VARIABLE MODELS

Authors: Maria-Pia Victoria-Feser
– University of Geneva,
Switzerland
maria-pia.victoriafeser@hec.unige.ch

Abstract:

- In this paper we study bias-corrections to the weighted MLE (Dupuis and Morgenthaler, 2002), a robust estimator simply defined through a weighted score function. Indeed, although the WMLE is relatively simple to compute, for most models it is not consistent and hence not very helpful. For example, the model we consider in this paper is the generalized linear latent variable model (GLLVM) proposed in Moustaki and Knott (2000) (see also Moustaki, 1996, Sammel, Ryan, and Legler, 1997 and Bartholomew and Knott, 1999). The score functions of this model are very complicated. They contain integrals that need to be evaluated. Moreover, they are highly nonlinear in the parameters which makes the use of complicated robust estimator quite impossible in practice. Moustaki and Victoria-Feser (2006) propose to use a weighted MLE and develop indirect inference (Gouriéroux, Monfort, and Renault, 1993, Gallant and Tauchen, 1996 and also Genton and de Luna, 2000, Genton and Ronchetti, 2003) to remove the bias. It can be computed in a simple iterative fashion. In this paper, we actually focus on indirect inference for bias correction in general. We rely heavily on the findings of Moustaki and Victoria-Feser (2006).

Key-Words:

- *factor analysis; latent variables; M-estimators.*

AMS Subject Classification:

- 62G35.

1. INTRODUCTION

Consider a general class of weighted MLE (WMLE) proposed by Dupuis and Morgenthaler (2002) belonging to the class of M-estimators (Huber, 1981) defined as the solution in $\boldsymbol{\theta}$ of

$$\frac{1}{n}\sum_{i=1}^{n}\psi_c(\boldsymbol{x}_i;\boldsymbol{\theta}) \;=\; \frac{1}{n}\sum_{i=1}^{n} s(\boldsymbol{x}_i;\boldsymbol{\theta})\, w(\boldsymbol{x}_i,c) \;=\; \mathbf{0}\;,$$

with the underlying assumption that $\boldsymbol{x}_i \sim F_{\boldsymbol{\theta}}$ and the weights $w(\boldsymbol{x}_i,c)$ are such that smaller weights are given to observations with larger score function $s(\boldsymbol{x}_i;\boldsymbol{\theta})= \partial/\partial\boldsymbol{\theta}\log\big(\partial/\partial\boldsymbol{x}\,F_{\boldsymbol{\theta}}(\boldsymbol{x})\big)$. A typical choice for the weights is Huber type weights, which for a given tuning parameter c are given by

$$w(\boldsymbol{x};c) \;=\; \min\left(1;\, \frac{c}{\|s(\boldsymbol{x};\boldsymbol{\theta})\|}\right), \tag{1.1}$$

where $\|...\|$ denotes the Euclidean norm. If $F_{\boldsymbol{\theta}}$ and/or the weight function are not symmetric, then the resulting estimator is not consistent. Based on a first order development of the bias, Dupuis and Morgenthaler (2002) propose a bias correction given by

$$B(\widehat{\boldsymbol{\theta}}) \;=\; -\left.\frac{\displaystyle\int s(\boldsymbol{x};\boldsymbol{\theta})\, w(\boldsymbol{x};\boldsymbol{\theta})\; dF_{\boldsymbol{\theta}}(\boldsymbol{x})}{\displaystyle\int\left(\frac{\partial}{\partial\boldsymbol{\theta}}\, s(\boldsymbol{x};\boldsymbol{\theta})\, w(\boldsymbol{x};\boldsymbol{\theta}) + s(\boldsymbol{x};\boldsymbol{\theta})\,\frac{\partial}{\partial\boldsymbol{\theta}}\, w(\boldsymbol{x};\boldsymbol{\theta})\right) dF_{\boldsymbol{\theta}}(\boldsymbol{x})}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \tag{1.2}$$

to be added to the inconsistent WMLE $\widehat{\boldsymbol{\theta}}$. The computation of two integrals is still needed (and can be done by means of simulations) as well as the derivative of the weight function. Alternatively, one can consider estimators of the type

$$\frac{1}{n}\sum_{i=1}^{n} s(\boldsymbol{x}_i;\boldsymbol{\theta})\, w(\boldsymbol{x}_i,c) - a(\boldsymbol{\theta}) \;=\; \mathbf{0}\;,$$

with

$$a(\boldsymbol{\theta}) \;=\int s(\boldsymbol{x};\boldsymbol{\theta})\, w(\boldsymbol{x},c)\; dF_{\boldsymbol{\theta}}(x)$$

and hence estimate simultaneously the bias correction with the estimator. This can become very complicated depending on the form of the score function. In the following section, a bias correction for a WMLE is presented, in the same spirit as (1.2) but based on the theory of indirect inference.

2. INDIRECT INFERENCE FOR BIAS REDUCTION

Indirect estimation (Gouriéroux, Monfort, and Renault, 1993, Gallant and Tauchen, 1996) has been proposed as an estimation procedure for a complex model $F_{\boldsymbol{\theta}}$ with intractable likelihood functions. It involves the computation of an estimator $\widehat{\boldsymbol{\pi}}$ for the parameters of an auxiliary model $F_{\boldsymbol{\pi}}$ that does not provide a consistent estimator of $\boldsymbol{\theta}$. In particular, let $\widehat{\boldsymbol{\pi}}$ be an M-estimator defined implicitly by

$$\int \psi(\boldsymbol{x};\widehat{\boldsymbol{\pi}})\, dF_n(\boldsymbol{x}) \;=\; \mathbf{0}\;.$$

As the sample size tends to infinity, this auxiliary estimate defines a mapping from the parameter space of $\boldsymbol{\theta}$ to the parameter space of the auxiliary model, i.e. $\boldsymbol{\theta} \to \boldsymbol{\pi}(\boldsymbol{\theta})$, defined by

$$\int \psi(\boldsymbol{x};\boldsymbol{\pi}(\boldsymbol{\theta}))\, dF_{\boldsymbol{\theta}}(\boldsymbol{x}) \;=\; \mathbf{0}\;. \tag{2.1}$$

With indirect inference one tries in some sense to invert this map, i.e. $\boldsymbol{\pi} \to \boldsymbol{\theta}(\boldsymbol{\pi})$, and use this inverse to calculate the estimator $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}(\widehat{\boldsymbol{\pi}})$. The latter can be obtained implicitly by the solution in $\boldsymbol{\theta}$ of

$$\int \psi(\boldsymbol{x};\widehat{\boldsymbol{\pi}})\, dF_{\boldsymbol{\theta}}(\boldsymbol{x}) \;=\; \mathbf{0}\;. \tag{2.2}$$

This indirect estimator results as a particular case of the general minimization problem defining indirect estimators, i.e.

$$\widehat{\boldsymbol{\theta}} \;=\; \arg\min_{\boldsymbol{\theta}} \big(\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}(\boldsymbol{\theta})\big)^T\, \Omega \big(\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}(\boldsymbol{\theta})\big)\;,$$

with $\boldsymbol{\pi}(\boldsymbol{\theta})$ obtained as the solution of (2.1). The matrix Ω can be chosen on efficiency arguments but for simplicity, on can choose $\Omega = I$. The estimation of $\boldsymbol{\pi}(\boldsymbol{\theta})$ is the difficulty in applying the indirect method. If it is possible to create samples $\widetilde{\boldsymbol{x}}_i(\boldsymbol{\theta})$, $i=1,...,s\cdot n$, simulated (with fixed seed) from $F_{\boldsymbol{\theta}}$ for a given $\boldsymbol{\theta}$, then a Monte Carlo estimate of $\boldsymbol{\pi}(\boldsymbol{\theta})$ can be used. This estimate is defined as the solution in $\boldsymbol{\pi}(\boldsymbol{\theta})$ of

$$\frac{1}{sn}\sum_{i=1}^{sn} \psi\big(\widetilde{\boldsymbol{x}}_i(\boldsymbol{\theta});\boldsymbol{\pi}(\boldsymbol{\theta})\big) \;=\; \mathbf{0}\;.$$

Gouriéroux, Monfort, and Renault (1993) show that this estimator is asymptotically equivalent to the one proposed by Gallant and Tauchen (1996) (available since 1992 as a working paper) defined by

$$\widehat{\boldsymbol{\theta}} \;=\; \arg\min_{\boldsymbol{\theta}} \left(\frac{1}{sn}\sum_{i=1}^{sn} \psi\big(\widetilde{\boldsymbol{x}}_i(\boldsymbol{\theta});\widehat{\boldsymbol{\pi}}\big)\right)^T \Delta \left(\frac{1}{sn}\sum_{i=1}^{sn} \psi\big(\widetilde{\boldsymbol{x}}_i(\boldsymbol{\theta});\widehat{\boldsymbol{\pi}}\big)\right), \tag{2.3}$$

with again Δ chosen on efficiency arguments. When $\dim(\boldsymbol{\theta}) = \dim(\boldsymbol{\pi})$ and $\Delta = I$, the solution of (2.3) is given by the solution in $\boldsymbol{\theta}$ of (2.2) in which the integral is estimated by the mean over a simulated sample. We also note that when ψ is the score function, then $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\pi}}$.

Indirect inference has already been used with robust statistics: see Genton and de Luna (2000) and Genton and Ronchetti (2003). Similar ideas can be found in Cabrera and Fernholz (1999).

$\widehat{\boldsymbol{\theta}}$ can be found iteratively using a Newton step. For that we need

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int \psi(\boldsymbol{x}; \widehat{\boldsymbol{\pi}}) \, dF_{\boldsymbol{\theta}}(\boldsymbol{x}) \; = \int \psi(\boldsymbol{x}; \widehat{\boldsymbol{\pi}}) \, s^T(\boldsymbol{x}; \boldsymbol{\theta}) \, dF_{\boldsymbol{\theta}}(\boldsymbol{x}) \; .$$

Then the Newton step is given by

$$\widehat{\boldsymbol{\theta}}^{(k+1)} \; = \; \widehat{\boldsymbol{\theta}}^{(k)} - S^{-1}\big(\widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\theta}}^{(k)}\big) \sum_{i=1}^{sn} \psi\big(\widetilde{\boldsymbol{x}}_i(\widehat{\boldsymbol{\theta}}^{(k)}); \widehat{\boldsymbol{\pi}}\big) \; , \tag{2.4}$$

where

$$S\big(\widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\theta}}\big) \; = \sum_{i=1}^{sn} \psi\big(\boldsymbol{x}_i(\widehat{\boldsymbol{\theta}}); \widehat{\boldsymbol{\pi}}\big) \, s^T\big(\boldsymbol{x}_i(\widehat{\boldsymbol{\theta}}); \widehat{\boldsymbol{\theta}}\big)$$

and $\widehat{\boldsymbol{\pi}}$ is the (inconsistent) M-estimator. With this indirect estimator, there is hence no need for simultaneous estimation of bias (computation of $a(\boldsymbol{\theta})$). This estimator has been proposed by Moustaki and Victoria-Feser (2006) in the context of generalized linear latent variable models.

3. GENERALIZED LINEAR LATENT VARIABLE MODELS (GLLVM)

Latent variable models are widely used in social sciences for studying the interrelationships among observed variables. More specifically, latent variable models are used for reducing the dimensionality of multivariate data, for assigning scores to sample members on the latent dimensions identified by the model as well as for the construction of measurement scales (e.g. in educational testing and psychometrics). Moustaki and Knott (2000) proposed a generalized linear latent variable model (GLLVM) framework for any type of observed data (metric and categorical) in the exponential family. They extended the work of Moustaki (1996) and Sammel, Ryan, and Legler (1997) for mixed binary and metric variables (the latter with covariate effects as well) and Bartholomew and Knott (1999) for categorical variables. A similar framework is also discussed by Skrondal and Rabe-Hesketh (2004) that includes multilevel models (random-effects models) as a special case.

Formally, given a set of response variables $x_1, ..., x_p$, there exists a (smaller) set of latent variables or factors $z_1, ..., z_q$ that account for the dependencies among the response variables. In other words, given the latent variables, the manifest ones are conditionally independent. Factor analysis is the simplest case. In general we suppose that the conditional distribution of the manifest variables given the latent ones belongs to the exponential family, i.e.

$$g_m\big(x_m \,|\, \boldsymbol{z}, \boldsymbol{\theta}_m\big) \; = \; \exp\left\{ \frac{x_m \, \boldsymbol{\alpha}_m \, \boldsymbol{z}^*}{\phi_m} - \frac{b\big(\boldsymbol{\alpha}_m \, \boldsymbol{z}^*\big)}{\phi_m} + c(x_m, \phi_m) \right\} , \qquad m = 1, ..., p \; ,$$

with $\boldsymbol{\alpha}_m = [\alpha_{m0}, ..., \alpha_{mq}]$, $m=1,...,p$, the so-called loadings, ϕ_m, $m=1,...,p$, the scale parameters (for example for normal manifest variables),

$$\boldsymbol{z}^* = [1, z_1, ..., z_q]^T = [1, \boldsymbol{z}^T]^T$$

and hence $\boldsymbol{\theta}_m = (\boldsymbol{\alpha}_m, \phi_m)^T$. The latent variables $\boldsymbol{z}$ are supposed standard multivariate normal with density $\varphi(\boldsymbol{z})$ (but the independence assumption can be relaxed), hence, the marginal distribution is

$$f(\boldsymbol{x};\boldsymbol{\theta}) = \int \cdots \int \left[\prod_{m=1}^{p} g_m(x_m | \boldsymbol{z}, \boldsymbol{\theta}_m) \right] \varphi(\boldsymbol{z}) \, d\boldsymbol{z} \, .$$

The score functions become

$$\begin{aligned} s_m^{(1)}(\boldsymbol{x};\boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\alpha}_m} \log\big(f(\boldsymbol{x};\boldsymbol{\theta})\big) \\ &= \frac{1}{f(\boldsymbol{x};\boldsymbol{\theta})} \int \cdots \int g(\boldsymbol{x} | \boldsymbol{z}, \boldsymbol{\theta}) \left(\frac{x_m - b'(\boldsymbol{\alpha}_m \boldsymbol{z}^*)}{\phi_m} \right) \boldsymbol{z}^* \varphi(\boldsymbol{z}) \, d\boldsymbol{z} \, , \end{aligned} \tag{3.1}$$

$$\begin{aligned} s_m^{(2)}(\boldsymbol{x};\boldsymbol{\theta}) &= \frac{\partial}{\partial \phi_m} \log\big(f(\boldsymbol{x};\boldsymbol{\theta})\big) \\ &= \frac{1}{f(\boldsymbol{x};\boldsymbol{\theta})} \int \cdots \int g(\boldsymbol{x} | \boldsymbol{z}, \boldsymbol{\theta}) \\ &\quad \cdot \left(- \frac{x_m \boldsymbol{\alpha}_m \boldsymbol{z}^* - b(\boldsymbol{\alpha}_m \boldsymbol{z}^*)}{\phi_m^2} + \frac{\partial}{\partial \phi_m} c(\phi_m, x_m) \right) \varphi(\boldsymbol{z}) \, d\boldsymbol{z} \, , \end{aligned} \tag{3.2}$$

for $m=1,...,p$. The integrals in (3.1) and (3.2) can be approximated using fixed Gauss–Hermite quadrature (see e.g. Bock and Liberman, 1970), adaptive quadrature points (see e.g. Bock and Schilling, 1997, Schilling and Bock, 2005), Monte Carlo approximations (see e.g. Sammel, Ryan, and Legler, 1997) or Laplace approximation (see e.g. Huber, Ronchetti, and Victoria-Feser 2004). All these approximations lead to approximate ML estimators. The models we consider here are one factor models and although it is known that Gauss–Hermite rule can give biased estimators in some situations, we will nevertheless use it to compute the integrals.

Moustaki and Victoria-Feser (2006) study the robustness properties of the (approximated) MLE by means of the Influence Function (Hampel, 1968, 1974). Not surprisingly, even with binary data, the MLE can be biased by data contamination, which in this context appear as unexpected binary responses. Since the (approximate) MLE is already quite complicate computationally, Moustaki and Victoria-Feser (2006) propose to use a WMLE with consistency correction via indirect inference. The WMLE $\widehat{\boldsymbol{\pi}}$ is computed with Huber type weights (1.1). The consistent estimator $\widehat{\boldsymbol{\theta}}$ is obtained using indirect inference and called Indirect Globally Weighted Robust (IGWR) estimator. Its (approximate) asymptotic covariance is also given in Moustaki and Victoria-Feser (2006) which is used for inference and also for choosing the tuning constant c of the Huber weights on efficiency arguments.

4. SIMULATION STUDY

We report here the simulation study presented in Moustaki and Victoria-Feser (2006). The model we consider is the one-factor model ($q=1$) fitted to two binary ($m=1,2$) and three normal ($m=3,4,5$) manifest variables with parameter values

- $\boldsymbol{\alpha}_1 = [1.0, 0.7]$,
- $\boldsymbol{\alpha}_2 = [0.8, 1.0]$,
- $\boldsymbol{\alpha}_3 = [2.0, 0.6]$ and $\phi_3 = 1$,
- $\boldsymbol{\alpha}_4 = [2.5, 0.7]$ and $\phi_4 = 1$,
- $\boldsymbol{\alpha}_5 = [3.0, 0.8]$ and $\phi_5 = 1$.

150 samples of size 200 where generated and contaminated in three ways:

- 3% of the first normal variable (i.e. observations of x_3) set to an arbitrary value (20) (pointmass 1);
- 3% of all three normal variables set to an arbitrary value (20) (pointmass 3);
- 3% of the data from the mixed GLLVM with $\boldsymbol{\alpha}_5 = [3.0, 8]^T$ instead of $\boldsymbol{\alpha}_5 = [3.0, 0.8]^T$ (model deviation).

The MLE, IGWR and IGWR1 which is defined by the iterative procedure given in (2.4) with only one iteration, were computed. The tuning constant was set to $c=3.5$, which corresponds to an efficiency level of 95% with respect to the MLE.

Figure 1 presents the distributions of the different estimators for the loading of the first manifest variable (binary) α_{11} with all types of contamination (including no contamination). Even if the contamination occurs on the normal manifest variables, the MLE can be biased as can bee seen with the pointmass 3 contamination type. Figures 2 and 3 present the distribution of the different estimators for respectively the mean of the third manifest variable (normal) α_{30} and the loading of the fifth manifest variable (normal) α_{51} with all types of contamination. The bias on the MLE appears quite large, while both robust estimators remain very stable. Without contamination, there is no apparent difference in distribution between the MLE and the robust estimators. Figure 4 presents the same analysis but for the estimators of the scale parameter for the first normal variable ϕ_3. The MLE of the scale parameter seems to be affected only when the contamination occurs only on the corresponding manifest variable. Again, the behavior of the robust estimators show great stability.

It should be noted that Moustaki and Victoria-Feser (2006) conclude that although the IGWR1 seems to perform very well with the examples of this simulations study, its bias increases more rapidly that the one of the IGWR as the WMLE is more biased, i.e. as the tuning constant c decreases.

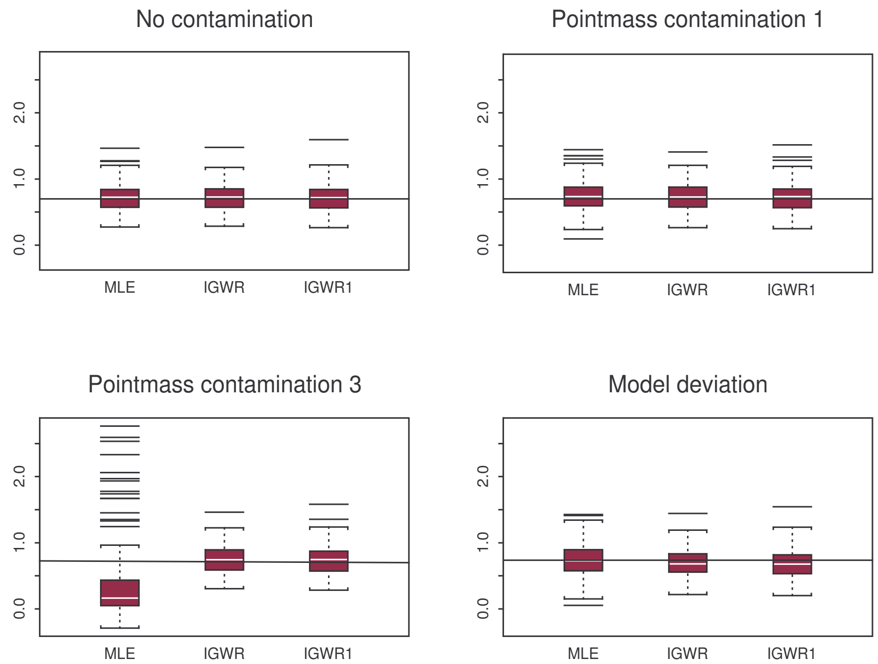

Figure 1: Distribution of the estimators for the loading on the first binary manifest variable. The horizontal line gives the true value.

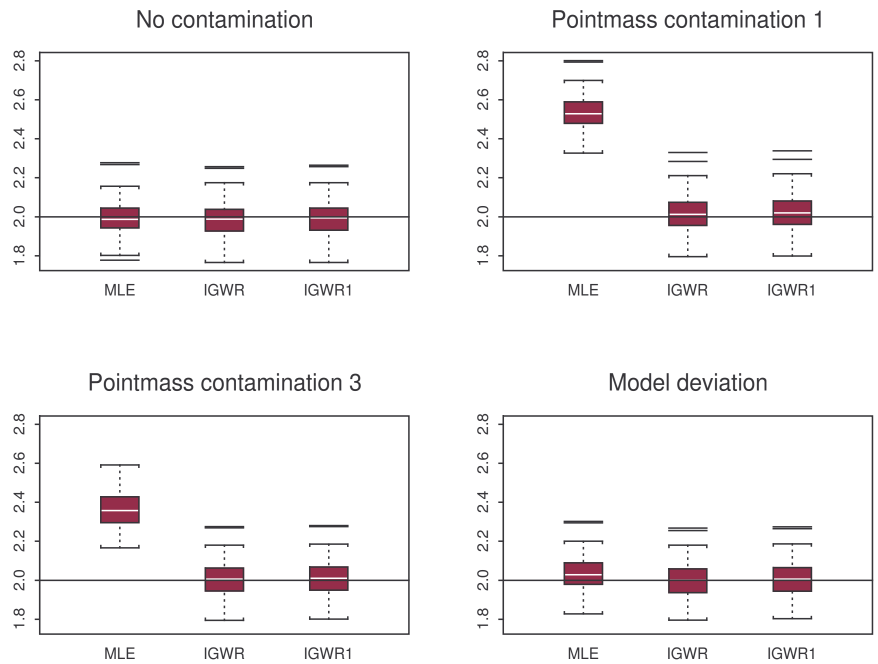

Figure 2: Distribution of the estimators for the mean on the first normal manifest variable. The horizontal line gives the true value.

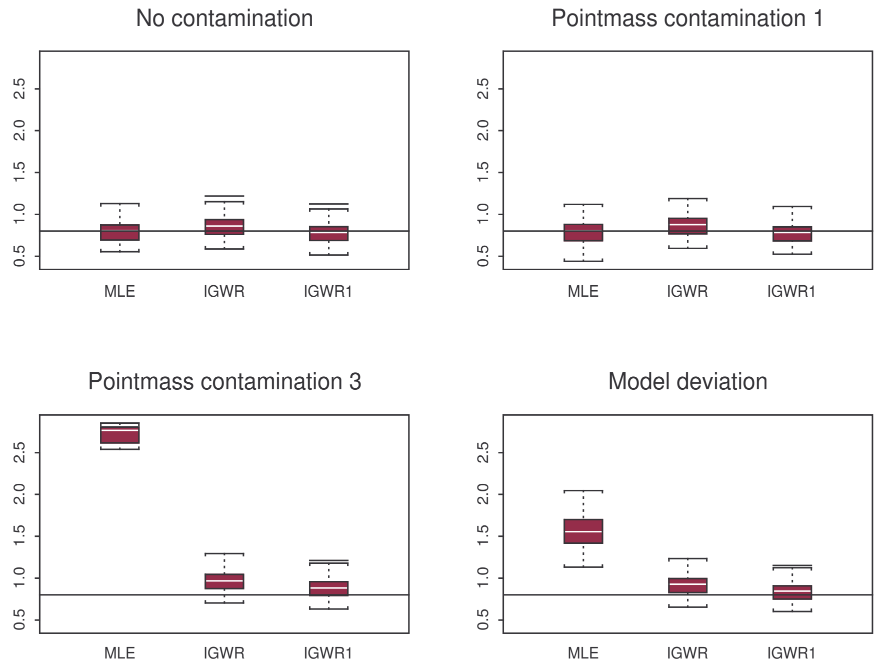

Figure 3: Distribution of the estimators for the loading on the third normal manifest variable. The horizontal line gives the true value.

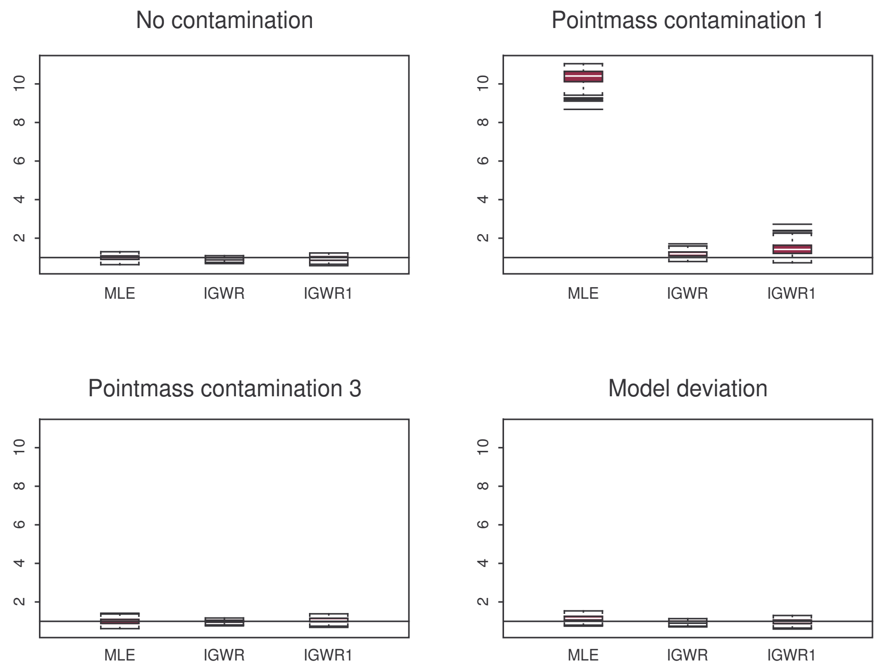

Figure 4: Distribution of the estimators for the scale on the first normal manifest variable. The horizontal line gives the true value.

5. ANALYSIS OF WEALTH DATA

Moustaki and Victoria-Feser (2006) also present an example based on a sub-sample of size 100 households of the 1990 consumption survey in Switzerland, provided by the Swiss Federal Statistical Office. The aim is to construct a measurement scale for the level of wealth, and for the purpose of this exercise, five variables have been selected. These are:

- purchase of a dishwasher (1/0) (Dishwasher)
- purchase of a car (1/0) (Car)
- equivalent food expenditure in logarithm (Food)
- equivalent expenditures for clothing in logarithm (Clothing)
- equivalent expenditures for housing in logarithm (Housing)

The continuous variables are treated as normal variables. Variables from the same survey have been analyzed before using the GLLVM by Moustaki and Knott (1997), Bartholomew and Knott (1999) and Huber, Ronchetti, and Victoria-Feser (2004). A one-factor model using both the ML and the IGWR estimators is fitted to the data. The bounding constant c has been set to 5 corresponding to an efficiency level of 94% (computed on the parameter values provided by the IGWR). The parameter values estimated by the ML and the IGWR estimators are presented in Table 1 together with their standard errors (the values in bold correspond to significant variables at the 5% level).

Table 1: Parameter's estimates and standard errors for the GLLVM on the wealth data.

Parameters		MLE		IGWR, $c = 5$	
		Estimate	Stand. Err.	Estimate	Stand. Err.
Constants	α_{10}	**−0.506**	0.23	**−0.589**	0.26
	α_{20}	**−0.623**	0.23	**−0.537**	0.23
	α_{30}	**6.922**	0.23	**6.887**	0.28
	α_{40}	**5.353**	0.32	**5.332**	0.32
	α_{50}	**7.087**	0.33	**7.140**	0.29
Loadings	α_{11}	0.466	0.26	**0.679**	0.28
	α_{21}	−0.167	0.24	0.216	0.25
	α_{31}	**1.021**	0.18	**1.098**	0.21
	α_{41}	**1.412**	0.31	**1.415**	0.28
	α_{51}	**1.044**	0.33	**1.064**	0.27
Variances	ϕ_3	0.289	0.16	**0.426**	0.17
	ϕ_4	**1.280**	0.27	**1.056**	0.20
	ϕ_5	**1.475**	0.22	**0.935**	0.14

The ML estimator shows that only the variables (Food, Clothing and Housing) are indicators of wealth, whereas the IGWR adds the variable Dishwasher. Both analyses exclude the variable Car. Variables Food and Housing are found with both methods to be indicators of the latent variable, whereas the association is stronger with the Clothing variable. For a diagnostics analysis, the weights given in (1.1) have been computed for each observation at the IGWR values and plotted in Figure 5. There are apparently (only) 5 outliers.

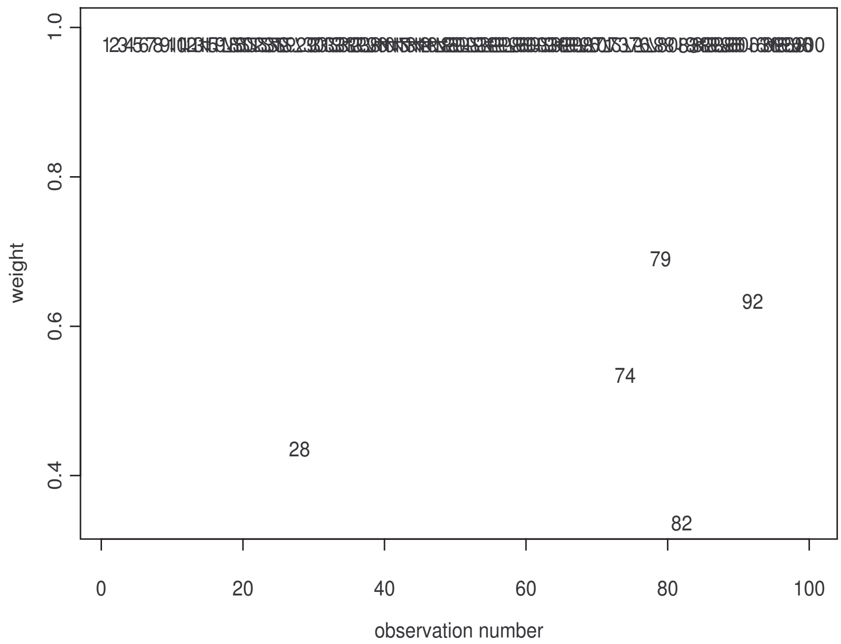

Figure 5: IGWR's weights against observation number for the wealth data.

ACKNOWLEDGMENTS

The author is partially supported by the Swiss National Science Foundation (grant #100012-109532).

REFERENCES

[1] Bartholomew, D.J. and Knott, M. (1999). *Latent Variable Models and Factor Analysis*, Arnold, London.

[2] Bock, R.D. and Liberman, M. (1970). Fitting a response model for n dichotomously scored items, *Psychometrika*, **35**, 179–197.

[3] Bock, R.D. and Schilling, S.G. (1997). *High-dimensional full-information item factor analysis*. In "Latent Variable Modelling and Applications to Causality" (M. Berkane, Ed.), Springer, New York, 164–176.

[4] Cabrera, J. and Fernholz, L.T. (1999). Target estimation for bias and mean square error reduction, *The Annals of Statistics*, **27**, 1080–1104.

[5] Dupuis, D.J. and Morgenthaler, S. (2002). Robust weighted likelihood estimators with an application to bivariate extremevalue problems, *Canadian Journal of Statistics*, **30**, 17–36.

[6] Gallant, A.R. and Tauchen, G. (1996). Which moments to match, *Econometric Theory*, **12**, 657–681.

[7] Genton, M.G. and de Luna, X. (2000). Robust simulation-based estimation, *Statistics and Probability Letters*, **48**, 253–259.

[8] Genton, M.G. and Ronchetti, E. (2003). Robust indirect inference, *Journal of the American Statistical Association*, **98**, 1–10.

[9] Gouriéroux, C.; Monfort, A. and Renault, A.E. (1993). Indirect inference, *Journal of Applied Econometrics*, **8** (supplement), S85–S118.

[10] Hampel, F.R. (1968). *Contribution to the Theory of Robust Estimation*, Ph.D. thesis, University of California, Berkeley.

[11] Hampel, F.R. (1974). The influence curve and its role in robust estimation, *Journal of the American Statistical Association*, **69**, 383–393.

[12] Huber, P.; Ronchetti, E. and Victoria-Feser, M.-P. (2004). Estimation of generalized latent trait models, *Journal of the Royal Statistical Society, Series B*, **66**, 893–908.

[13] Huber, P.J. (1981). *Robust Statistics*, John Wiley, New York.

[14] Moustaki, I. (1996). A latent trait and a latent class model for mixed observed variables, *British Journal of Mathematical and Statistical Psychology*, **49**, 313–334.

[15] Moustaki, I. and Knott, M. (1997). *Generalized latent trait models*, Statistics research report 36, London School of Economics.

[16] Moustaki, I. and Knott, M. (2000). Generalized latent trait models, *Psychometrika*, **65**, 391–411.

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

[18] Sammel, M.D.; Ryan, L.M. and Legler, J.M. (1997). Latent variable models for mixed discrete and continuous outcomes, *Journal of the Royal Statistical Society, Series B*, **59**, 667–678.

[19] Schilling, S. and Bock, D.R. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature, *Psychometrika*, **70**, 533–555.

[20] Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*, Chapman and Hall, London.