DataFirst Technical Papers

DataFirst SALDRU
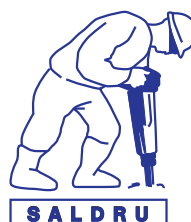
Sample Survey Calibration: An Informationtheoretic perspective

by
Martin Wittenberg

Recommended citation

Wittenberg, M. (2009). Sample Survey Calibration: An Information theoretic perspective. A DataFirst Technical Paper Number 14. Cape Town: DataFirst, University of Cape Town

Sample Survey Calibration: An Information-theoretic perspective

Martin Wittenberg
School of Economics and SALDRU
University of Cape Town

June 2009

Abstract

We show that the pseudo empirical maximum likelihood estimator can be recast as a calibration estimator. The process of estimating the probabilities p_k of the distribution function can be done also in a maximum entropy framework. We suggest that a minimum cross-entropy estimator has attractive theoretical properties. A Monte Carlo simulation suggests that this estimator outperforms the PEMLE and the Horvitz-Thompson estimator.

Keywords: sample weights, calibration, pseudo-empirical maximum likelihood estimation, cross entropy

Most statistical agencies calibrate their surveys to external benchmarks in order to increase the precision of the estimates. A common approach is to minimise a distance measure between the design weights and the calibrated weights while ensuring that the calibrated weights satisfy the benchmark requirements (Deville and Särndal 1992, Deville, Särndal and Sautory 1993). Deville and Särndal (1992) have argued that such "calibration estimators" are asymptotically equivalent to generalized regression estimators and that "numerical features of the weights and ease of computation become more than anything else the bases for choosing between the estimators" (p.376). A more theoretical approach to the choice of estimator is given by Chen and Sitter (1999). They argue that their "pseudo empirical maximum likelihood estimator" (PEMLE) is not only attractive conceptually, but is also likely to be efficient when compared to other alternatives. In this paper we will argue that the minimisation of a cross-entropy criterion provides an alternative approach which is theoretically coherent and performs at least as well as the PEMLE in a set of Monte Carlo simulations. Interestingly it provides theoretical justification for iterative "raking ratio" adjustments. We show, furthermore, that it allows for easy generalisation to cases other than dummy variables. In addition it allows for the straightforward incorporation of constraints at different levels, such as households and individuals.

1   Calibration estimators

The theory of calibration estimators is developed in Deville and Särndal (1992). Let d_k be the design-weight associated with unit k, i.e. $d_k = \frac{1}{\pi_k}$ where π_k is the inclusion probability, i.e. $\pi_k = \Pr(k \in s)$ where s is the sample. A distance measure G between the weights d_k and the "calibrated" weights w_k has to satisfy the following criteria:

1. $G(w_k, d_k) \geq 0$; G is differentiable with respect to w_k; strictly convex; and $G(d_k, d_k) = 0$

2. $\frac{\partial G}{\partial w_k} = g(w_k, d_k)$ is continuous and one-to-one.

The average distance between the vectors $\mathbf{w}$ and $\mathbf{d}$ will be estimated by

$$\sum d_k G(w_k, d_k) \tag{1}$$

We assume that we have auxiliary information about the population total of the vector of random variables $\mathbf{x}_k$ i.e. $\sum_{k \in U} \mathbf{x}_k = \mathbf{t_x}$, where U is the universe. We want the calibrated weights to return this total. Consequently the objective is:

$$\min_{\mathbf{w}} \sum_{k \in s} d_k G(w_k, d_k), \text{ s.t. } \sum_{k \in s} w_k \mathbf{x}_k = \mathbf{t_x}$$

Deville and Särndal (1992) pick G so that it varies only with the ratio $\frac{w_k}{d_k}$, i.e. $G\left(w_k, d_k\right) \equiv G\left(\frac{w_k}{d_k}\right)$. Forming the Lagrangian for this problem, differentiating it and setting the derivative equal to zero we get

$$g\left(\frac{w_k}{d_k}\right) - \mathbf{x}_k' \boldsymbol{\lambda} = 0$$

Letting $F\left(\cdot\right) = g^{-1}\left(\cdot\right)$ we get the solution

$$w_k = d_k F\left(\mathbf{x}_k' \boldsymbol{\lambda}\right) \qquad (2)$$

The Lagrange multipliers $\boldsymbol{\lambda}$ can be obtained by solving the "calibration equation"

$$
\begin{aligned}
\mathbf{t_x} &= \sum_{k \in s} w_k \mathbf{x}_k \\
\mathbf{t_x} &= \sum_{k \in s} d_k F\left(\mathbf{x}_k' \boldsymbol{\lambda}\right) \mathbf{x}_k
\end{aligned}
\qquad (3)
$$

Different choices of G lead to different calibration estimators. Deville and Särndal (1992, p.378) list *inter alia* some of the following possibilities:

Table 1: Examples of distance functions used for calibration

Case		$d_k G\left(w_k, d_k\right)$	$F\left(\mathbf{x}_k' \boldsymbol{\lambda}\right)$
1	Linear – Generalized regression estimator	$\left(w_k - d_k\right)^2 / 2d_k$	$1 + \mathbf{x}_k' \boldsymbol{\lambda}$
2	Multiplicative	$w_k \log\left(\frac{w_k}{d_k}\right) - w_k + d_k$	$\exp\left(\mathbf{x}_k' \boldsymbol{\lambda}\right)$
3	"Minimum entropy distance"	$-d_k \log\left(w_k/d_k\right) + w_k - d_k$	$\left(1 - \mathbf{x}_k' \boldsymbol{\lambda}\right)^{-1}$

Note that Deville and Särndal's function G_k is equivalen to $d_k G$.
Our notation here is consistent with Deville et al. (1993)

Deville and Särndal (1992, p.378) note that the first choice can some times yield negative weights, which can be problematic. The multiplicative model can yield extreme weights, which can be unpalatable to end users. The third case, they suggest, need not always have a solution, although the probability of a solution will approach one as $n \to \infty$. They prove that all these calibration estimators are asymptotically equivalent to the generalised regression estimator, hence their claim that the choice of an estimator can be based on pragmatic grounds. To that end they develop estimators that can restrict the range of the calibrated weights.

2 Pseudo Empirical Maximum Likelihood Estimation

A different approach is adopted by Chen and Sitter (1999).who suggest that a design-consistent estimate of the population empirical likelihood function $l\left(F\right)$ is given by

$$\widehat{l}\left(\mathbf{p}\right) = \sum_{k \in s} d_k \log\left(p_k\right) \qquad (4)$$

Any finite population parameter θ_N of the finite population distribution function F_N or superpopulation parameter θ can therefore be estimated in two steps. Firstly $\widehat{l}\left(\mathbf{p}\right)$ is maximised to obtain $\widehat{\mathbf{p}}$ subject to the constraint $\sum_{k \in s} p_k = 1$. An estimate of the finite population distribution function is then given by

$$\widehat{F}_N\left(x\right) = \sum_{k \in s} \widehat{p}_k I_{[x \le x_k]} \qquad (5)$$

where $I_{[.]}$ is the indicator function. In the second step $\widehat{\theta}_N$ is obtained as a function of $\widehat{F}_N$. For instance if $\theta_N = E_{F_N}\left(x\right)$, i.e. the finite population mean, then

$$\widehat{\theta}_N = \sum_{k \in s} \widehat{p}_k x_k$$

If we have auxiliary information about any of the finite population moments, e.g. $E(x) = \overline{x}$, then we maximise the sample empirical likelihood (equation 4) subject to the constraints $\sum_{k \in s} p_k \mathbf{x}_k = \overline{\mathbf{x}}$ which can also be written as

$$\sum_{k \in s} p_k (\mathbf{x}_k - \overline{\mathbf{x}}) = \mathbf{0} \tag{6}$$

This problem can again be solved through Langrange multipliers. The solution is given by

$$p_k = q_k \left(1 + \boldsymbol{\lambda}' (\mathbf{x}_k - \overline{\mathbf{x}})\right)^{-1} \tag{7}$$

where $q_k = \frac{d_k}{\sum d_k}$ and $\boldsymbol{\lambda}$ solves the equations

$$\sum_{k \in s} \frac{q_k (\mathbf{x}_k - \overline{\mathbf{x}})}{\left(1 + \boldsymbol{\lambda}' (\mathbf{x}_k - \overline{\mathbf{x}})\right)} = \mathbf{0} \tag{8}$$

Comparing these solutions to equations 2 and 3 and the solution to case 3 of Table 1, it is evident that the PEMLE is, in fact, equivalent to the "minimum entropy distance" calibration estimator of Deville and Särndal. The only difference is that the estimator returns $\widehat{p}_k$ rather than w_k. This is simply a normalisation issue: with

$$w_k = N\widehat{p}_k \tag{9}$$

we get a set of calibrated weights.

Let us add the constraint $\sum_{k \in s} w_k = N$, which is equivalent to the constraint $\sum_{k \in s} p_k = 1$ and rewrite the constraints $\mathbf{t_x} = \sum_{k \in s} w_k \mathbf{x}_k$ as

$$N^{-1}\mathbf{t_x} = \sum_{k \in s} p_k \mathbf{x}_k$$

The constraints in terms of totals are obviously equivalent to constraints in terms of means. It is easy to see now that.the two approaches are mathematically identical, since

$$\sum \left(-d_k \log \frac{w_k}{d_k} + w_k - d_k\right) = \sum -d_k \log w_k + \sum d_k \log d_k + N - \sum d_k$$
$$= \sum -d_k \log N p_k + \sum d_k \log d_k + N - \sum d_k$$

is minimised at precisely the same values as $\sum d_k \log p_k$ is maximised.

3 Minimum cross-entropy estimation

The innovative part of the PEMLE procedure is to view the process of calibration as a process of estimating the underlying distribution function (in equation 5). Viewed in this way the process can be seen as trying to solve an "ill-posed problem" (Golan, Judge and Miller 1996), i.e. one where the number of unknowns (p_k) exceed the information at hand (the number of constraints). Maximum entropy estimation is designed precisely for these sort of cases. We would solve the problem

$$\max_{\mathbf{p}} \sum_{k \in s} p_k \log p_k, \text{ s.t. } \sum_{k \in s} p_k \mathbf{x}_k = \overline{\mathbf{x}} \text{ and } \sum_{k \in s} p_k = 1 \tag{10}$$

Comparing this to the PEMLE, we see that the objective function can be thought of as $E(\log p_k)$ where the expectation is taken with respect to the distribution F, while the empirical log-likelihood can be thought of as $N \cdot E(\log p_k)$ where the expectation is taken with respect to different possible samples.

The maximum entropy estimator ignores the sampling information contained in the design weights d_k. An alternative approach is to think of the design weights as providing a prior set of estimates $q_k = \frac{d_k}{\sum d_k}$ of the probabilities to be estimated. The problem is therefore to pick a probability vector $\mathbf{p}$ that is as close as possible

to $\mathbf{q}$ while respecting the moment constraints. The approach to this problem is given by the "cross-entropy formalism" (Golan et al. 1996, pp.29ff). This can be written as

$$\min_{\mathbf{p}} \sum_{k \in s} p_k \log \frac{p_k}{q_k}, \text{ s.t.} \sum_{k \in s} p_k \mathbf{x}_k = \overline{\mathbf{x}} \text{ and } \sum_{k \in s} p_k = 1 \tag{11}$$

This problem yields the solution

$$\widehat{p}_k = q_k \exp\left(\mathbf{x}_k' \widehat{\boldsymbol{\lambda}}\right) / \Omega, \qquad \Omega = \sum_{k \in s} q_k \exp\left(\mathbf{x}_k' \widehat{\boldsymbol{\lambda}}\right) \tag{12}$$

As with the PEMLE, the minimum cross-entropy estimator (MCEE) can yield raising weights by simply multiplying the probabilities by N, as in equation 1. In this case the MCEE turns out to be mathematically equivalent to Deville and Särndal's second case, the multiplicative calibration estimator[1]. Asymptotically they are therefore equivalent to each other and equivalent to the generalized regression estimator.

There are nevertheless some attractive features of the minimum cross-entropy approach which are not shared by the other approaches. Firstly, the MCEE is based on trying to minimise the additional information required in moving from the prior distribution to the distribution $\mathbf{p}$. High information in this context means that there are only a few states of the world which satisfy the constraints. Minimising this additional information is therefore equivalent to picking a distribution which can give rise to the observed information (the finite population means $\overline{\mathbf{x}}$) with higher probability than alternative ones. The MCEE therefore has a theoretical rationale which some of the ad-hoc calibration estimators considered by Deville and Särndal do not.

Secondly, the cross-entropy criterion obeys the law of composition (the importance of this in the maximum entropy context is stressed by Jaynes 1957). Let $\mathbf{I}(\mathbf{p}; \mathbf{q})$ be the cross-entropy measure, i.e.

$$\mathbf{I}(p_1, p_2, \dots, p_n; q_1, q_2, \dots, q_n) = \sum_{k=1}^{n} p_k \log \frac{p_k}{q_k} \tag{13}$$

Now assume that the data was collected in two stages: households and individuals within households. Let p_{ih} be the probability that individual i in household h was sampled, $p_{\cdot h} = \sum_i p_{ih}$ is the probability of observing household h and $p_{i|h} = p_{ih}/p_{\cdot h}$ is the probability of observing individual i, given that household h was sampled. Similarly we define $q_{\cdot h}$ and $q_{i|h}$. The cross-entropy measure satisfies the following relationship:

$$\begin{aligned} \mathbf{I}(p_{11}, \dots, p_{h_n n}; q_{11}, \dots, q_{h_n n;}) &= \mathbf{I}(p_{\cdot 1}, \dots, p_{\cdot n}; q_{\cdot 1}, \dots, q_{\cdot n}) + p_{\cdot 1} \mathbf{I}(p_{1|1}, \dots, p_{h_1|1}; q_{1|1} \dots, q_{h_1|1}) \\ &\quad + p_{\cdot 2} \mathbf{I}(p_{1|2}, \dots, p_{h_2|2}; q_{1|2} \dots, q_{h_2|2}) + \dots + p_{\cdot n} \mathbf{I}(p_{1|n}, \dots, p_{h_n|n}; q_{1n} \dots, q_{h_n|n}) \end{aligned} \tag{14}$$

In many cases it makes sense to assume that the conditional probabilities $p_{i|h}$ and $q_{i|h}$ are equal. For instance if all household members are enumerated if the household is selected we have $p_{i|j} = q_{i|j} = \frac{1}{h_j}$. In these cases $\mathbf{I}(p_{1|j}, \dots, p_{h_j|j}; q_{1|j} \dots, q_{h_j|j}) = 0$. Consequently

$$\mathbf{I}(p_{11}, \dots, p_{h_n n}; q_{11}, \dots, q_{h_n n;}) = \mathbf{I}(p_{\cdot 1}, \dots, p_{\cdot n}; q_{\cdot 1}, \dots, q_{\cdot n})$$

Furthermore the constraint $\sum_{k \in s} p_k \mathbf{x}_k = \overline{\mathbf{x}}$ can now be rewritten as

$$\begin{aligned} \sum_h \sum_i p_{ih} \mathbf{x}_{ih} &= \sum_h \sum_i p_{\cdot h} p_{i|h} \mathbf{x}_{ih} \\ &= \sum_h p_{\cdot h} \sum_i q_{i|h} \mathbf{x}_{ih} \\ &= \sum_h p_{\cdot h} \overline{\mathbf{x}}_h \end{aligned}$$

---

[1] Merz and Stolze (2008) calibrate data using a "Minimum Information Loss" criterion. Their approach is identical to that of the MCEE except that they do not interpret it as estimating probabilities. They merely wish to minimise the distance between w_k and d_k. They do not seem to see that this is equivalent to the calibration estimator.

where $\overline{\mathbf{x}}_h$ is the estimated mean of $\mathbf{x}$ within household h. Consequently minimising the cross-entropy across individuals in terms of equation 11 while imposing the condition $p_{i|h} = q_{i|h}$ is equivalent to the problem:

$$\min_{\mathbf{p}.} \sum_{h \in s} p_{\cdot h} \log \frac{p_{\cdot h}}{q_{\cdot h}}, \text{ s.t. } \sum_{h \in s} p_{\cdot h} \overline{\mathbf{x}}_h = \overline{\mathbf{x}} \text{ and } \sum_{h \in s} p_{\cdot h} = 1$$

which is a household level minimum cross-entropy problem. We can therefore incorporate constraints about the intra-household relative weights in a straightforward manner.

4   Comparing the PEMLE and MCEE: a Monte Carlo experiment

Both the PEMLE and MCEE are guaranteed to produce positive weights. They both have cogent theoretical rationales and they are asymptotically equal to the generalized regression estimator and thus to each other. The MCEE has the advantage that it is designed specifically for the first stage estimation problem, i.e. to recover the probabilities p_k which allow us to estimate the distribution function F_N. Secondly it obeys the composition law, so that we might expect it to perform better in estimating conditional probabilities or in incorporating constraints at different levels.

Nevertheless it is unclear on purely theoretical grounds how these estimators might perform in finite samples. In order to investigate this matter, we ran a Monte Carlo experiment as follows: We constructed 200 "censuses" of 5000 observations each. In each census there were two "strata". Units were randomly allocated to the strata according to whether a pseudorandom number was larger than $0.65 + u$ where u was a draw from the uniform distribution with support $(-0.2, 0.2)$, fixed for each census. Each census therefore had a different balance between the strata. Within each of the strata a random log-normal "income" variable x was generated. In the smaller stratum the distribution was $LN(8 + \mu, 1)$ where $\mu \in (0, 0.4)$, fixed for each census, while in the larger stratum the distribution was $LN(7 + \mu, 1)$. This income variable was, in turn, coded into 11 income bands.

Once each census had been fixed in these ways, 200 stratified random samples without replacement were extracted from each census, a sample of size 50 from the high income stratum ($s = 1$) and a sample of size 50 from the low income stratum ($s = 0$). Given the known stratum sizes, the standard Horvitz and Thompson (1952) weights could be calculated and these in turn were fed as prior weights d_k to the PEMLE and MCEE. We also calibrated the PEMLE and MCEE to the census stratum sizes and the census mean income. Within each census we could estimate the bias and mean square error resulting from the weights in the calculation of various statistics. In particular we focussed on the conditional mean of income within each stratum plus the estimated proportions in each income band. The latter exercise was designed to test how accurately the PEMLE and MCEE were able to estimate a crude approximation to the distribution function F_N. In all cases we compared the estimated statistic $\widehat{\theta}_N$ to the finite population statistic θ_N calculated over the census.

The performance of the estimators across the 200 censuses is given in Table 2. In the first three columns we report the average bias of the estimators. With the exception of the MCEE's bias on the top income bracket $\widehat{p}_{b=11}$ none of these bias estimates is significantly different from zero at the 95% level, i.e. if the bias estimates from the 200 censuses are ranked from the most negative to the most positive, then zero is inside the interval defined by the range from the 5th value to the 195th one.

The next three columns report the average mean square error of the estimators and the last two columns report a nonparametric 95% confidence interval for the difference between $\widehat{\theta}_{MCEE}$ and $\widehat{\theta}_{PEMLE}$. This is calculated from $MSE\left(\widehat{\theta}_{MCEE}\right) - MSE\left(\widehat{\theta}_{PEMLE}\right)$ where this difference is calculated across the same 200 samples for a given census. For instance the first row shows that across all 200 censuses the MCEE had **on average** a smaller mean square error than the PEMLE. The nonparametric confidence interval shows that in over 195 of the censuses the gap in the MSE ranged from -89793.3 to -145.3. This is reasonable evidence that the MCEE is more accurate in estimating the conditional mean than the PEMLE is. Indeed the table shows that the MCEE is more accurate on average in the calculation of most of the statistics and where it is not the difference is not statistically significant. By contrast for many of the statistics the MCEE outperforms the PEMLE in at least 95% of the censuses.

Table 2: Performance of different weights in a Monte Carlo experiment

	Bias			MSE			95% CI for difference	
	HT	MCEE	PEMLE	HT	MCEE	PEMLE	5th	195th
$E(x\mid s=1)$	7.673	24.902	41.481	1216396.0	522448.9	539981.0	-89793.3	-145.3
$E(x\mid s=0)$	-1.239	-0.432	-7.456	163761.6	109629.7	112520.0	-9181.3	-179.6
$\widehat{p}_{b=1}$	0.000006	-0.000035	-0.000009	0.000044	0.000043	0.000044	-0.000002	0.000000
$\widehat{p}_{b=2}$	0.000113	-0.000090	0.000033	0.000215	0.000210	0.000213	-0.000007	-0.000001
$\widehat{p}_{b=3}$	0.000225	-0.000466	-0.000048	0.000697	0.000672	0.000679	-0.000017	-0.000001
$\widehat{p}_{b=4}$	-0.000305	-0.001743	-0.000877	0.001446	0.001385	0.001393	-0.000025	0.000008
$\widehat{p}_{b=5}$	0.000069	-0.001693	-0.000740	0.001994	0.001910	0.001915	-0.000030	0.000016
$\widehat{p}_{b=6}$	-0.000204	-0.000904	-0.000834	0.002013	0.001995	0.001993	-0.000013	0.000027
$\widehat{p}_{b=7}$	0.000149	0.001476	0.000144	0.001499	0.001528	0.001495	-0.000002	0.000072
$\widehat{p}_{b=8}$	-0.000152	0.002529	0.000958	0.000776	0.000800	0.000771	-0.000012	0.000106
$\widehat{p}_{b=9}$	0.000027	0.001696	0.001761	0.000286	0.000256	0.000269	-0.000034	-0.000002
$\widehat{p}_{b=10}$	0.000039	-0.000241	0.000066	0.000070	0.000059	0.000064	-0.000012	-0.000002
$\widehat{p}_{b=11}$	0.000033	-0.000527	-0.000454	0.000013	0.000007	0.000008	-0.000001	0.000000

HT - Horvitz Thompson, PEMLE - pseudo-empirical maximum likelihood estimator

MCEE - minimum cross-entropy estimator. Bias $= \widehat{\theta}_N - \theta_N$, MSE$=(\widehat{\theta}_N - \theta_N)^2$

Results based on 200 "Censuses" of 5000 observations each, from each of which 200 stratified random samples of size 100 were extracted

5   Conclusion

The MCEE is attractive for theoretical reasons and it seems to work well in finite samples, at least for the case considered by our Monte Carlo experiment. This raises the question why this "multiplicative" calibration estimator is not used more frequently by statistical agencies. The reason is probably due to the fact that it can generate extreme weights, as emphasized by Deville and Särndal (1992, p.378). Equation 12 shows that $\widehat{p}_k$ will be large relative to q_k only if $\mathbf{x}'_k\boldsymbol{\lambda}$ is large. A large Lagrange multiplier implies that that particular constraint is very informative, i.e. it reduces radically the uncertainty about the underlying probability distribution. As such it should be seen not as a problem with the technique but a problem deriving from imposing constraints on the sample that it is ill-equipped to deal with. Disguising this problem by limiting the acceptable range of weights does not improve the quality of the information derived from the use of such calibrated weights.

References

Chen, Jiahua and R.R. Sitter, "A Pseudo Empirical Likelihood Approach to the Effective Use of Auxiliary Information in Complex Surveys," *Statistica Sinica*, 1999, *9*, 385–406.

Deville, Jean-Claude and Carl-Erik Särndal, "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, 1992, *87* (418), 376–382.

___ , ___ , and Olivier Sautory, "Generalized Raking Procedures in Survey Sampling," *Journal of the American Statistical Association*, 1993, *88* (423), 1013–1020.

Golan, Amos, George Judge, and Douglas Miller, *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, Chichester: Wiley, 1996.

Horvitz, D.G. and D.J. Thompson, "A Generalization of Sampling Without Replacement From a Finite Universe," *Journal of the American Statistical Association*, 1952, *47* (260), 663–685.

Jaynes, E.T., "Information Theory and Statistical Mechanics," *Physical Review*, 1957, *106* (4), 620–630.

Merz, Joachim and Henning Stolze, "Representative time use data and new harmonised calibration of the American Heritage Time Use Data (AHTUD) 1965-1999," *electronic International Journal of Time Use Research*, 2008, *5* (1), 90–126.

About DatatFirst

DataFirst is a research unit at the University of Cape Town engaged in promoting the long term preservation and reuse of data from African Socioeconomic surveys. This includes:

• the development and use of appropriate software for data curation to support the use of data for purposes beyond those of initial survey projects
• liaison with data producers - governments and research institutions - for the provision of data for reanalysis
• research to improve the quality of African survey data
• training of African data managers for better data curation on the continent
• training of data users to advance quantitative skills in the region.

The above strategies support a well-resourced research-policy interface in South Africa, where data reuse by policy analysts in academia serves to refine inputs to government planning.

DataFirst

www.datafirst.uct.ac.za
Level 3, School of Economics Building, Middle Campus, University of Cape Town
Private Bag, Rondebosch 7701, Cape Town, South Africa
Tel: +27 (0)21 650 5708