ESSA Conference Paper, 3-5 September 2019, Johannesburg

Restoring Representativeness to South African Household Survey Data with Cross-Entropy Weight Recalibration

Amy Thornton[1*] and Martin Wittenberg[1]

[1]School of Economics, University of Cape Town, South Africa
[*]Corresponding author: amy.thornton@uct.ac.za

Abstract

Calibration is the process of fine-tuning household survey weights so that estimates conform with known population totals, sourced from an auxiliary data base. This process improves the extent to which the sample drawn represents the actual population, thereby improving the quality of inferences. Calibration is essentially an optimisation problem, meaning the quality of the calibration is therefore partly determined by modelling decisions on the part of the survey statistician, such as choice of functional form and number of restrictions. If done poorly, calibration can create more problems that it solves, as was the case with this paper's case study of South African national survey data. In this case, a sub-optimal calibration technique introduced unnecessary noise into estimation and set up an incoherent and unrealistic conceptual framework. A consequence of the resultant weighting scheme was a needless curtailing of the number of questions the data could be used to answer. Since this process was embedded in more than 20 years worth of data, these issues represent a significant efficiency cost to the public funds spent on its collection. Fortunately, these deficiencies can be overcome. In this paper, we employ cross-entropy estimation to recalibrate the survey weights for a stacked series of cross-sections from the South African October Household Survey and General Household Survey between 1994-2011. The recalibration is successful in resolving the inconsistencies introduced by the existing calibration process and so measurably improves the quality of the data. More generally, the study demonstrates how existing data resources can be improved.

Keywords: calibrated weights, weight recalibration, cross-entropy, survey calibration

National household survey data collected on a frequent basis by country statistics bureaus are key resources for policy and research agendas. Many of the questions researchers and policy-makers have cannot be answered by any other type of data, e.g. the annual collection of labour market statistics or national-level poverty monitoring. Researchers are reliant on the strength of these data in order to produce high-quality research in these fields. The quality of the output feeds not only into the quality of the research, but into the quality of the information and conclusions contributed to policy discussions. Data quality of national household surveys is therefore an important topic.

One aspect of data quality is how well the data at hand represents the sampling frame, which in the case of national survey data, is the country population. If the data is insufficiently representative, this limits the types of questions the data can be used to answer as well as the generalisability of research conclusions. The consequence is a diminishment of the public investment in the collection of national household survey data which is a very costly exercise usually happening on an annual - if not more frequent - basis.

The degree to which data is representative depends on factors at every point in the data collection and dissemination process, such as sampling design, fieldworker practice, and post-sampling calibration of survey weights. Calibration of survey weights is one technique commonly used by survey statisticians to improve the representativeness of a survey. Calibration is the process of weighting the sample so that estimates agree with known population totals, extracted from a census, for example. By making the sample 'look more like' the population, the quality of inferences is improved.

Although calibration is used to fix deficiencies in representativeness, if done poorly, it can create more problems. This was the case in South African national survey data which provides a useful case study of exactly how calibration can go awry and compromise the data quality of an otherwise well-designed survey. More importantly, though, the South African case also demonstrates how faulty calibration can be fixed. The nature of the calibration process as well as the fact that it is undertaken after a survey is collected means it can be remedied relatively painlessly compared to other influences on data quality, like survey design or sampling errors.

In the South African data, information was collected at both the person and household level, making both of these elements important units of analysis. Curating a consistent and reliable series of population and household counts over time was, therefore, a first-order output of the survey data. Difficulty arose, though, in calibrating the sample to agree with auxiliary demographic projections for the person count and the household count at the same time. To overcome this problem, calibration for these two main units of analysis was simply separated into two parallel and mutually exclusive calibration models, yielding a separate person and household weight.

This dual weighting scheme introduces unnecessary noise into inference and the resulting reduction in representativeness needlessly limits the number of research questions this data can be used to answer in a conceptually coherent way. Since this procedure is embedded in more than 20 years worth of data, such an erosion of data quality counts as a considerable cost both to the country's research agenda and to the efficiency of public spending.

Fortunately, this cost is not incontrovertible. In this paper, we are able to use cross-entropy estimation to recalibrate the survey weights for a stacked series of cross-sections between 1994-2011. The new cross-entropy weight coherently links people to the households they live in, thereby restoring the mutual representativeness of people and households to the data. Additionally, the new weight yields reliable and consistent demographic estimates and recovers the number of questions these data can be coherently employed to answer. Our research shows that with careful thought and consideration we can improve the quality of existing data resources.

The next section introduces the data and outlines exactly why representativeness was impeded in the process of trying to square a household stock with a person population. Section 3 details the cross-entropy technique and how we are able to reconcile people and households into a single weight. Section 4 presents our results, both in terms of the validity of the calibration itself, as well as, how our new weight solves the problems created by the current practice. We present caveats on the scope of our improvements to the weighting system, before Section 5 concludes.

1 Setting Up Quality Household Survey Data

The best type of survey data for collecting information about people and households is a census since this surveys the entire population. However, censuses are very expensive operations and entail large-scale complicated logistics so that they are not usually feasible on a regular-enough basis for the type of monitoring household surveys are often used for. This is especially the case in developing countries with capacity constraints on their statistics bureaus. It is much more feasible to survey a sample of the population. This smaller operation is more cost effective, can be carried out more regularly, and allows for more detailed and directed questionnaires. Crucially, with a combination of prudent survey design and careful survey weighting, statistics estimated from the sample will satisfactorily approximate what they would have been had they been estimated from a census (Deaton, 1997). Ensuring that the sample is representative of the population, or sampling frame, has received a lot of scholarly attention in recent decades and is the topic of an extensive academic and practical literature (Deaton, 1997; Lavallée and Beaumont, 2015; Deville and Särndal, 1992).

Survey weights are an essential aspect of representativeness. Survey weights are scaling factors assigned to each unit of observation to make the sample representative of the population. There are usually three steps to this process: (1) assigning the design weights which account for probability of a unit being selected for sampling, in turn dependent on survey design; (2) adjusting the weights to compensate for units that were selected to be surveyed but were not surveyed for some reason (e.g. non-response; non-coverage); and (3) tuning the weights so that estimates from the weighted sample reflect known population totals, often taken from a census, using a process known as calibration (Lavallée and Beaumont, 2015).

Common survey designs, such as two-stage sampling and stratification, usually mean the probability that a given unit will be selected for the sample varies from unit to unit (Deaton, 1997). This means that each unit in the sample represents a different number of units in the population and thus the sample needs to be weighted accordingly to achieve unbiased estimates of population means. This initial weighting is the role of the design weights, which are defined as the inverse of the probability of a unit being selected for sampling, and, when summed, will yield the population total. Most national surveys collect information at the household and person level, which means that the probability that a person is selected for sampling is conditional on the probability that their household is selected for sampling.

Weights additionally need to be adjusted for accidental ex-post survey factors (Lavallée and Beaumont, 2015). Invariably, the actual sample collected differs from that which was designed due participant non-response, non-coverage, or other measurement error. As such, the probability of inclusion will vary across sample units for reasons other than design. The survey statistician models non-response in order to make this adjustment. Non-response adjustment is less rigorous than the design weight since modelling decisions by the survey statistican are embedded in the weight.

The third stage of weighting, which this paper is most concerned with, is calibration. Calibration is the process of weighting the sample so that the estimates conform with known population totals from an auxiliary data base, often demographic characteristics about age, gender and geography (Deaton, 1997).[1] The statistical problem of calibration is that unnecessary randomness is introduced into estimation by the process of sampling and sampling design. The aim is to improve the quality of inferences in case a poor sample is selected by weighting the sample to make it as representative as possible of the true population.

Key work in this field comes from Deville (2000) and Deville and Särndal (1992) who present how calibrated weights, w_i, can be solved for by minimising a distance function between estimated population totals using the design weights and known population totals from an auxiliary data source. Let x_i be the value of a variable for the ith observation in a sample s of size n drawn from population size N. The population total for $x = (x_1, ..., x_i, ..., x_N)$ is defined as $t_x = \sum_{i=1}^{N} x_i$. Consider now, that we know the value of t_x from some auxiliary information source like a census. An unbiased estimator of t_x would involve weighting up the observation values with the design weights, d_i, and then summing so that $\sum_{i \in s} d_i x_i = \hat{t}_{x,d}$.

If a bad sample is selected, this estimator could be very far from the actual t_x in which case calibration is motivated. The aim is to choose new calibrated weights, w_i, to be as close to the design weights as possible in order to maintain unbiasedness (Deville and Särndal, 1992), but also so that they agree with the known population totals so that:

$$\hat{t}_{x,w} = \sum_{i \in s} w_i x_i = t_x \tag{1}$$

Since the new weights should be kept as similar as possible to the design weights, the optimisation problem is therefore to minimise a distance function $G_i(w_i, d_i)$. That is, to minimise $\sum_{i \in s} G_i(w_i, d_i)$ subject to equation 1 (Särndal, 2010). This yields the following solution for the calibrated weights:

$$w_i = d_i F(q_i, x_i, \lambda) \tag{2}$$

where $1/q_i > 0$ is a set of known scaling factors[2] unrelated to d_i; and, λ is solved using the constraint in equation 1 and has a probability tending to one as $n \to \infty$.

The calibration is ultimately informed by a whole vector of known population totals of a size less than the total number of observations. A basic example is that a census could inform us about the number of women in the population, the number of people living in rural areas, as well as the number of people aged 25-35. The calibration therefore needs to simultaneously optimise the weights for gender, region, and age, putting several restrictions on rural women aged 25-35.

As such, obtaining the calibrated weights is not as objective a process as obtaining the design weights which depend entirely on the survey design (Deaton, 1997). By contrast, calibration is

[1] The design weights will not always conform with population totals because in some cases, the survey is designed to over- or under-sample certain sub-populations for cost or statistical efficiency reasons. For example, in South Africa, the Indian sub-population is very small. As such, they are usually oversampled in national household surveys in order to collect a large enough sample for robust estimation. This means that they are designed to occur in the sample at a higher rate than what they occur in the population.

[2] The scaling factors arise from the choice of $G(w, d)$. The function is chosen so that the first partial derivative with respect to w is $g(w, d) = \partial G(w, d)/\partial w = g(w/d)/q$.

essentially a modelling problem and the modelling can be of better or worse quality, depending on the standard of the auxiliary information available, as well as, the modelling decisions made by the survey statistician. There are many different ways to specify the distance function that needs to be optimised ($G_i(w_i, d_i)$) as well as the scaling factors, q_i (Särndal, 2010). An array of different approaches are documented in an extensive literature on the topic of the theory of calibration (Särndal, 2010).

The number of known population totals, or restrictions, to use is also an important choice. On the one hand, the more restrictions that are used (e.g. age, region, race) and the more detailed (e.g. finer age bands, more disaggregated regions), the closer the sample should get to resembling the true population and the more precise your estimates. However, too many restrictions relative to raw sample size can undermine the computing of the calibration as well as introduce other small sample biases as the sample is cut into finer and finer cells (Deville and Särndal, 1992). The survey statistician thus faces a trade-off between the improvement in the accuracy of a core set of estimates and the potential distortion caused by the calibration. For example, ensuring accurate person and household counts may be prioritised, at the cost of misshaping another, less fundamental statistic.

Modelling decisions by the survey statistician are thus embedded in calibrated weights and researchers may be more or less in agreement with these decisions and may want to make their own choices (Deaton, 1997). For these reasons, it is important that calibration method is well-documented and that design weights are released along with calibrated weights. Finally, note that calibration can also be undermined if the external source providing the known population totals is of poor quality or the sampling errors from non-coverage or non-response are large and difficult to overcome. In these cases, calibration can create as many problems as it tries to solve (Deaton, 1997).

South African national household survey data provides an interesting case study of how survey weight calibration can undermine data quality and how these issues can be overcome with a good understanding of weighting techniques. In the South African case, a sub-optimal calibration approach resulted in a very restrictive and unrealistic conceptual framework being embedded into almost 20 years worth of data and introduced unnecessary noise into estimation. The data sets were also not released with the original design weights, meaning researchers were unable to make their own decisions about how to proceed with weighting and instead had to adopt the faulty weights provided. The next section introduces the South African data and explains the calibration technique, as well as, the problems it creates.

2 Case Study: Survey Weight Calibration in South African National Household Survey Data

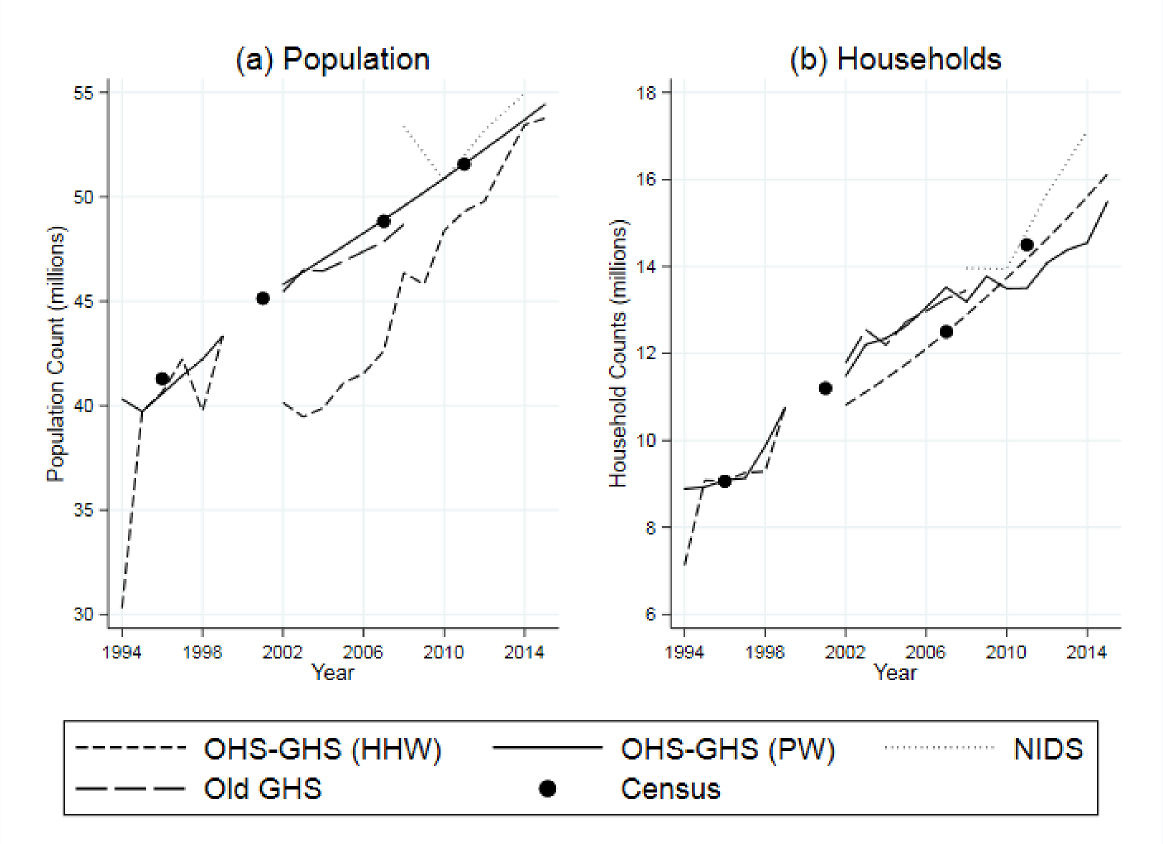
The two most fundamental demographic statistics found in any national household survey data are, perhaps, the number of people and the number of households. Because they are so foundational, if the data cannot accurately represent these two counts, confidence in any analysis on more complex statistics is undermined. These two statistics are often the frame to which survey weight calibration is tethered and - if the survey data is collected on a regular basis - attention is dedicated to achieving a consistent series of these counts. For these reasons, person and household counts are discussed in this section as a sign of the health of the survey data.

As previously discussed, the best tool for counting people and households in a country is a census. South Africa has census data for 1996, 2001, and, 2011; as well as, a Community Survey in 2007.[3] Outside of these years, we rely on annually collected household survey data. Census and various household survey data sets yield the trends in Figure 1 describing the number of people and households in South Africa over the post-apartheid period. The census counted just over 41 million people living

[3] Community Surveys are larger-than-usual national household surveys conducted in lieu of a census due to capacity constraints (Statistics South Africa, 2007*b*). The 2007 Community Survey had a sample size of 949 thousand people and 246 thousand households, making it about three times the size other South African household national household surveys (Statistics South Africa, 2007*a*).

in 9 million households in 1996; in 2011, there were just over 51 million people living in 14.5 million households.

Figure 1: Total Population and Household Counts in South Africa in Various Data Sources, 1994-2015



Source: own calculations.
Notes: OHS-GHS = Stacked series of the October Household Survey (1993 - 1999) and the General Household Survey (2002 - 2015); HHW = Household weighted; PW = Person weighted; NIDS = National Income Dynamics Study; Old GHS = GHS weighted using the original integrated weight; Census = Census 1996, 2001, and 2011 and Community Survey in 2007.

Out of the set of household survey data sets collected by South Africa's national statistics bureau, Statistics South Africa (SSA), the most appropriate to use for counting people and households is the October Household Survey (OHS) and the General Household Survey (GHS). The OHS is the only large nationally representative annual household survey undertaken by SSA in the period 1993-1999. Additionally, the OHS collected data about relationships within the households, defined in reference to the household head (e.g. spouse of the head, child of the head). After 1999, the OHS was conceptually split into the Labour Force Surveys and General Household Surveys, with the former focusing on economic outcomes and the latter on socio-demographic outcomes. The GHS - which only launched in 2002 - is therefore the survey that inherited questions about household relationships from the OHS, making it the logical choice for the latter half of the period.

Both the OHS and the GHS are cross-sectional and survey approximately 30 000 dwelling units based on about 3 000 Primary Sampling Units drawn from the Master Sample of enumerator areas used during the most recent census at the time. Exceptions are that the 1996 and 1998 October Household Surveys only surveyed about 16 000 and 20 000 dwelling units, respectively. A stratified, two-stage cluster sampling design is employed in each case, stratified at the provincial level.[4] Data is self-reported to the enumerator (or by proxy in the case of an absent respondent) and covers the spectrum from demographic and household information to basic labour market data. Between these two surveys then, we have large samples of nationally representative cross-sectional data on individuals, households, and their structures for every year in the period 1993-present, with the exception of 2000 and 2001.

As mentioned initially, household surveys only sample a portion of the population. The danger with this approach as opposed to a census is that the sample collected can end up looking quite different to the actual population because of non-response, luck surrounding the drawing of a single random sample, and other measurement error occurring during the surveying process. Consequently, weights in household surveys are calibrated as previously described so that the sample resembles the population as a whole. However, there are some important causes for concern about the quality of calibration in these series, and in particular, in the GHS.

In the OHS and the GHS, sampling practice is that if a household is sampled, all people living in that household are sampled. The implication is that the household weight and the person weight for people living in that household should be equal. This rule has not always been consistently applied. For the OHS, SSA calibrated the person weights in such a way that people within a household had different weights (Branson and Wittenberg, 2014). This raised the problem of what the household weight should be and resulted in the release of a separate household weight (and for most years of the OHS it is unclear how the final household weight was calibrated.[5]) Having two different and incoherent person and household weights means that in some cases researchers will arrive at two different estimates of the same statistic, even though they are using the same set of data. An example of this outcome can be viewed in the discrepancy between the OHS-GHS (HHW) and (PW) series in both panels of Figure 1 in the OHS period, 1993-1999.

From 2002 and with the beginning of the GHS series, SSA used integrated weights. That is, they ensured that people within the same household had the same weight; and then used this as the household weight. Although this was consistent with sampling practice, concerns were raised when these weights yielded unrealistic trends in the number of households. An uneven trend came out over the period 2002-2008 (Statistics South Africa, 2010), observable in the Old GHS series plotted in Panel (b) of Figure 1. This was possibly a consequence of the weights only being calibrated on person and not household information.

To overcome the difficulty of simultaneously squaring a person and household population, the household and person weight calibration was instead separated again from 2008 to present. The person weights continue to be integrated weights, equal within the household and calibrated only on person information. The household weight, though, is now calculated completely separately using a headship model that uses household-level information (Statistics South Africa, 2010). Like in the OHS, having different person and household weights leads to multiple inconsistent estimates. The problem in the case of the GHS is even more pronounced as the two weights are completely divorced from each other due to their calibration models being based on different input information. The inconsistency this introduces into statistic estimation is again evident from the disagreement between the OHS-GHS (HHW) series and the OHS-GHS (PW) series in both panels of Figure 1 in the GHS period, 2002-2015. In Panel (a), the person weight yields a smooth, evenly-increasing population series, but the household weight produces a highly divergent and unstable series for the same statistic. The opposite is true for the case of households in Panel (b). This is evidence that the person and household universe are not reliably represented concurrently using either weight. Researchers must choose between ‘two worlds’: one in which people are consistently represented, but not the households they live in, or, vice versa.

[4] The 2004 Master Sample was stratified at the district council level.

[5] For example, metadata from years 1997-9 describes designing the household weight, but only post-stratifying the person weight (Statistics South Africa, 1997, 1998, 1999). In 1994, though, the weight of the household head was used as the household weight (Wittenberg, 2008).

The result of this compartmentalising of representativeness has consequences beyond estimation. By only allowing either people or households to be accurately represented at a time, any analysis that crosses the person-household boundary (e.g. per capita household income welfare analysis) is rendered incoherent in a conceptual sense. This happens because the conceptual framework of the data is set up so that people and households are treated as existing independently and in isolation of one another. The implication is that person behaviour plays out independent of our groupings into households; households are understood to have behaviour patterns of their own, divorced from the individuals that comprise them who instead are treated as behaving as a single homogeneous unit. If this sounds unrealistic, it is because it is intuitively and logically at odds with how social scientists think about people and households. Consider, for instance, the notion that the household you live in has no bearing whatsoever on your welfare outcomes. The upshot is that researchers can only perform either person-level or household-level analysis in isolation, if they want to be conceptually coherent, thus seriously restricting the types of questions these data can be used to answer.

To make matters worse, this problem is not restricted to the GHS alone, but is present in all SSA household and labour market survey data. The Income and Expenditure Survey (IES), the Living Conditions Survey (LCS), and the GHS all have a separate person and household weights. The Labour Force Surveys (LFS), Quarterly Labour Force Surveys (QLFS), and Labour Market Dynamics Survey (LMDS) series are also affected. SSA only provide a person weight for the (Q)LFS and LMDS series. This implies that researchers - probably unknowingly - endorse the faulty conceptual framework associated with this weighting system. Any household-level analysis carried out in the labour market surveys is therefore questionable, since person weights are not calibrated using household-level information and SSA themselves found this problem enough to calibrate an entirely new weight just for household-level analysis.

Ultimately, the dual-weight system serves not only to introduce noise and inconsistency into estimates of statistics, but also to cement into the foundations of the data an unrealistic conceptual framework that divorces people from the households they live in. The implication of the dual-weight system is that unless researchers are doing in-sample analysis only or using their own more coherent weighting scheme, many research conclusions from SSA data sets since 2008 (the abandonment of the combined household and person weight) are corrupted as described and consequently so are any government policy or planning discussions based on these conclusions. The immediate importance of these data sets for economic planning cannot be understated. For example, the IES is used to compute South Africa's Consumer Price Index and estimate poverty lines and the (Q)LFS series is the benchmark for labour market monitoring. The inconsistency and incoherence of the dual-weight system has therefore permeated research and policy outcomes in South Africa quite comprehensively.

South African household survey data quality has been compromised by a reduction in representativeness. This issue stems from a sub-optimal survey weight calibration strategy which separately calibrated weights for people and households, forcing the weighting system to break with sampling practise. The survey design, however, allows for the concurrent representation of people and households. An alternative approach, then, is to combine the models into a single cohesive one yielding a single calibrated weight, consistent with sampling practise. In the next sections we put forward our alternative calibration technique, cross-entropy estimation, and describe how we applied it in the South African case. The section following that discusses how, in addition to achieving a sound calibration, our recalibration overcomes the specific shortcomings of the SSA dual-weight system: the inconsistency introduced into estimation of statistics and the incoherence of the conceptual framework.

3 Recalibrating Survey Weights using Cross-Entropy Estimation

Our goal is to recalibrate the survey weights for the full data series of the OHS and GHS spanning 1994-2015 using cross-entropy estimation in order to improve the degree to which the sample represents the population. An immediate deliverable of this goal is achieving more consistent and conceptually coherent person and household counts over the post-apartheid period. The outcome will be a single weight that assigns the same weight to everybody living in the same household, as well as, to the household, itself. An important contribution here is having both person and household information from the censuses enter into the cross-entropy estimation as a constraint. In other words, this weight will consolidate all the same information SSA uses to calibrate their separate weights into a single weight. This then clearly and logically links people to the households they live in, expanding the conceptual framework in a sensible way, and restoring representativeness to the data.

The generalities of the calibration problem are essentially the same as those laid out in Section 1, with the cross-entropy technique using an entropy measure as its distance function. Given moment constraints and a normalisation restriction, cross-entropy minimises the information ($\mathbf{I}(\mathbf{p}, \mathbf{q})$) needed to make the new distribution of weights ($\mathbf{p}$) resemble as much as possible information we already have about what it should look like, which could be a distribution of existing SSA calibrated or design weights ($\mathbf{q}$). This can be formalised as:

$$\mathbf{I}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{q_i}\right) \tag{3}$$

where $\mathbf{I}(\mathbf{p}, \mathbf{q})$ is minimised subject to:

$$y_j = \sum_{i=1}^{n} X_{ji} p_i, j = 1, ..., J \tag{4}$$

$$\sum_{i=1}^{n} p_i = 1 \tag{5}$$

where there are J population moments; y_i is the population mean of the random variable X_j; and equation 5 is the normalisation restriction. This constrained optimisation problem can be solved by maximising the unconstrained dual cross-entropy objective function:

$$L(\lambda) = \sum_{j=1}^{J} \lambda_j y_j - ln[\Omega(\lambda)] = \mathbf{M}(\lambda) \tag{6}$$

where $\Omega(\lambda)$ is given by:

$$\Omega(\tilde{\lambda}) = \sum_{i=1}^{n} q_i \exp(\mathbf{x}_i \tilde{\lambda}) \tag{7}$$

Golan et al. (1997) show that this function behaves much like maximum likelihood. The function $\mathbf{M}$ can be characterised as an expected log likelihood where $\mathbf{p}(\lambda)$ is the exponent and the parameter is λ. The parameter λ reveals the extent to which the new distribution, $\mathbf{p}$, is a distortion of the original underlying distribution, $\mathbf{q}$. Each constraint in the constraint matrix has an associated λ coefficient indicative of how informative that particular constraint is in estimating $\mathbf{p}$. A higher λ coefficient, for example, is an indication that the respective constraint resulted in the $\mathbf{p}$ distribution moving relatively further away from the original $\mathbf{q}$ distribution. For more detail, the reader is referred to Wittenberg (2010) and Golan et al. (1997).

In other words, the technique weights the sample according to information about what the sample looks like, was designed to look like, as well as, information about what the population looks like. We use the *maxentropy*.ado package in STATA 15 to carry out the estimation. The distribution of prior information, **q**, comes from the design weights in the case of the GHS, and the calibrated weights in the case of the OHS. Population moments enter the command via a constraint matrix which is set up based on auxiliary information from SSA.

The sample for each year is obtained from the publicly accessible data set downloaded from South African data repository DataFirst's website. Design weights are not publicly released for either the OHS or the GHS. The design weights for the GHS period 2002-2011 were received in private correspondence between Martin Wittenberg and SSA. As a result of only having weights for this period, we only run the recalibration up until 2011. We have to use SSA's calibrated weights in the case of the OHS, and these were available in the publicly available data sets.[6]

The constraint matrix includes both individual-level as well as household-level moments, a contribution of this paper. The individual-level population moments come from SSA's Mid-Year Population Estimates (MYPE), which are publicly available for the period 2002 to present on the SSA website (i.e. the GHS period). Population estimates for the OHS period came from back projections of the population from DataFirst (2018). This was necessary as SSA do not provide disaggregated enough population estimates for my purposes for these years.[7] The household moments come from household headship rates calculated by SSA for their household weight model from the three available censuses (1996, 2001 and 2011), as well as, the 2007 Community Survey. Headship rates are interpolated for intervening years.

The final constraint matrix consists of 103 constraints: 63 individual age-sex-race dummies; 8 province dummies; and, 32 age-sex-race household headship dummies.[8] The calibration compiled successfully for all years in the interval 1994 - 2011. These data sets are then appended to each of other after cleaning and harmonising a selected set of variables to achieve a complete and coherently weighted data set that can represent both people and households simultaneously in the period 1994 - 2011.

4 Comparing the Cross-Entropy Weight to the Existing Dual-Weight Scheme

There are two problems with the dual-weight system: inconsistency introduced into estimation and an incoherent implicit conceptual framework of people and households. In order to show that the cross-entropy weight is a useful alternative to the dual-weight system, it needs to perform better on both these counts. Since there is only one weight for people and households, we already know that the cross-entropy weight won't be producing inconsistent estimates; by which we mean, there will only be one estimate per statistic. Instead, what is worth evaluating, is the soundness of the calibration itself. This involves assessing the degree to which the new weight is a distortion of the prior weighting distribution (the **q** distribution) and whether it is able to reproduce its own constraints (the census and MYPE). This is covered in Section 4.1. In Section 4.2, we evaluate the coherence of the conceptual foundation of each weighting system - SSA dual-weight and cross-entropy - by the degree to which each system is internally consistent.

[6] Although we are still in correspondence with SSA, there are no design weights still on record for the OHS to our knowledge.

[7] The population was back-projected by applying an exponential model of population growth to SSA MYPE from 2002 onwards and using population growth rates from the Actuarial Society of South Africa (ASSA).

[8] The 64th individual and 9th province dummy are excluded as these sets are mutually exclusive. There were eight age categories for the individual dummies in intervals of ten years, beginning with 0 - 9 years and ending with those aged 80 years and above. There were four age categories for the household headship dummies. These were 0 - 34 years; 35 - 49 years; 50 - 64 years; and 64 years and older.

4.1 Evaluating the Validity of the Cross-Entropy Calibration

We begin by comparing summary statistics of the cross-entropy weight (CEW) with summary statistics from the prior $\mathbf{q}$ distribution in Table 1. The $\mathbf{q}$ distribution came from the person weight in the case of the OHS and the design weight in the case of the GHS. Table 1 reports the range of the two weight distributions in each year of the series, as well as some diagnostics about the λ coefficient output which is a signal about the health of the calibration. The λ coefficient is a measure of the extent to which the prior distribution is distorted by the constraints to yield the CEW distribution. High λ coefficients are therefore indicative of constraints that have resulted in a CEW distribution that is substantially different to the prior weight distribution. There is no particular rule regarding how high a λ is too high, but five is a useful rule of thumb (Wittenberg, 2010), reported in the last column.

The CEW tends to range higher than the prior distribution, especially in the OHS period. The year that stands out the most is 1994 which has a minimum CEW of zero[9] and an enormous maximum of over 200 thousand. The 1994 calibration is the weakest based on the λ coefficient diagnostics: about a third of coefficients are higher in absolute terms than five. It should be noted, though, that the 1994 sample is unusual for being different in important ways to the sample used for the rest of the series. In 1994 Whites were oversampled and Africans were undersampled. In a sense then, high λ coefficients in 1994 are indicative of constraints that are highly informative not so much in distorting the sample, as 'distorting it back' to what it should look like.

Some of the highest λ coefficients came out for older, usually male, age groups for both the individual and household headship categories. The exception was a high degree of adjustment required for the shares of African women over 70 years in the GHS period in particular. In the OHS period, Asian/Indian and Coloured men over 80 years needed more adjustment. The headship constraints in general were more informative than the individual and provincial ones and this was particularly the case for white male heads over 50 years. These patterns are likely reflective of smaller samples of older people, especially males.

The calibration settled down from 1995 onwards. The shares of absolute λ coefficients that exceed five are negligible after 1994 and - aside from 1995 - the average absolute λ is always less than two. The 2005 calibration is noteworthy because it has an especially high maximum, but the rest of the λ coefficient diagnostics do not draw attention. Noting the weakness in 1994, we conclude that the calibration is satisfactorily similar to the prior distribution.

Figure 2 comprises the main result displaying the quality of the cross-entropy weight (CEW) in estimating population and household counts in comparison to the SSA person (PW) and household (HHW) weight. For reference, these figures are reported in table format in Table 2. All three weights are trying to match the estimates from those of the census or MYPE. The CEW outperforms both the PW and HHW by this criterion. In Panel (a), the CEW line lies exactly on top of the census and MYPE estimates of the population for the full period 1994 - 2011. The PW matches the MYPE exactly in the GHS period, but slightly undercounts the population in the OHS period compared to the CEW.[10] The HHW, on the other hand, produces a jagged trend line that appears quite unrelated to the census/MYPE population trend. Further to this, Table 2 records that the CEW is able to match the MYPE and census counts at the 1 million level (with the exception of 1998). That the CEW is able to reproduce the census or MYPE so well demonstrates the success of the calibration since these estimates were used as constraints.

Turning to household counts, there is a minor discrepancy between the CEW household count and the census count in Table 2. The largest margin of error occurs in 1996: the census counted

[9] The zero CEW weights in 1994, 1997, and 1998 are indicative of missing data required for the calibration. In addition to seven cases of zero-valued CEWs, there are 79 cases of missing CEWs. In the case of the seven zeroes, these occurred in 1997 and 1998 when the PW was missing completely - the weight used as the prior distribution for the CEW. There were three missing CEWs in 1996 and 76 in 1998, all of which coincide with missing PWs and sometimes also household heads.

[10] Note that the MYPE for the OHS period do not come from SSA directly as detailed in the figure notes.

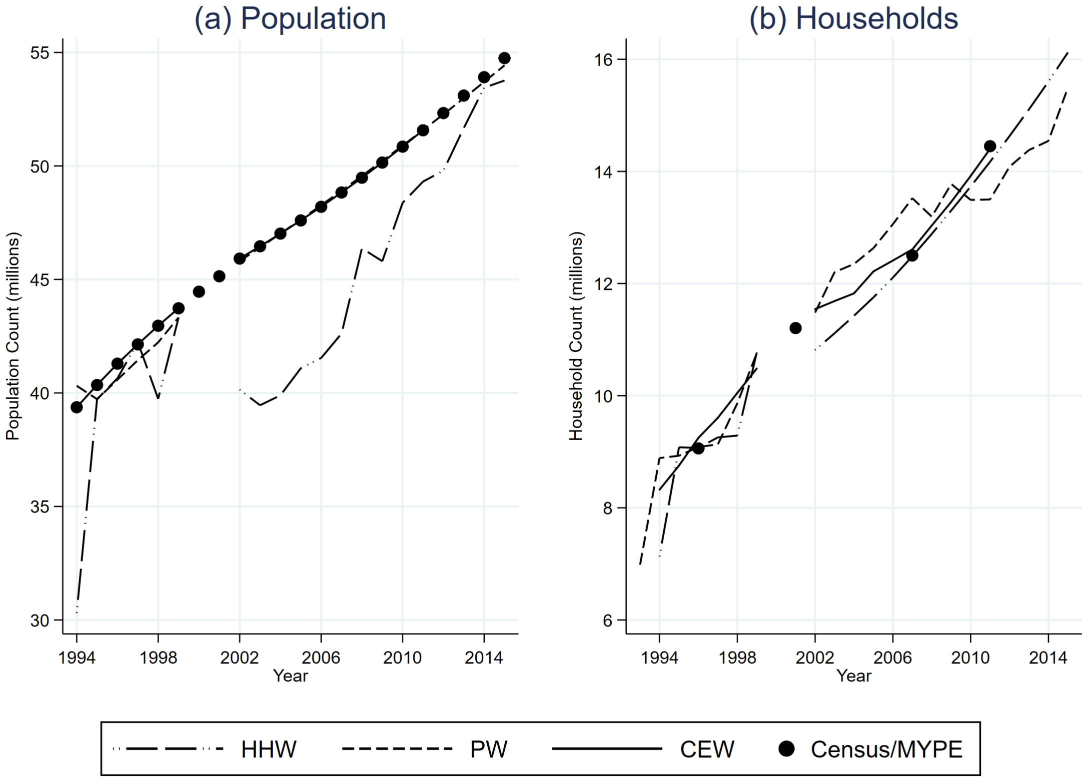
Table 1: Summary and Diagnostic Statistics for the Prior Weight Distribution and the Cross-Entropy Weight (CEW) in a Stacked Series of the OHS-GHS

	Prior Distribution		**CEW**		**Lamda (λ) Diagnostics**		
							Share of
Year	**Min.**	**Max.**	**Min.**	**Max.**	**| Mean |**	**| Max. |**	$\lvert\lambda\rvert\geq$**5**
1994	14.43	1 936.71	0.00	206 570.90	4.10	13.21	0.32
1995	0.07	1 759.65	0.12	6 134.33	2.22	8.48	0.07
1996	61.77	6 053.41	11.74	6 432.31	1.57	6.91	0.03
1997	42.00	1 834.00	0.00	5 506.44	1.87	5.78	0.03
1998	47.66	2 629.73	0.00	7 432.17	1.10	4.82	0.00
1999	12.71	2 387.87	13.13	9 522.72	1.45	5.68	0.03
2000							
2001							
2002	16.38	8 437.75	6.73	9 384.52	1.78	5.85	0.02
2003	6.35	7 986.57	2.50	12 096.30	1.82	6.86	0.02
2004	8.04	10 264.53	4.88	10 962.53	1.37	5.05	0.01
2005	5.90	5 969.12	3.74	20 056.11	1.19	3.95	0.00
2006	5.89	11 435.05	1.45	10 053.11	1.41	4.47	0.00
2007	5.89	7 277.31	1.19	11 472.63	1.12	4.18	0.00
2008	115.43	7 275.32	10.95	9 061.39	0.94	3.89	0.00
2009	115.43	5 171.03	10.84	8 536.67	1.46	5.02	0.01
2010	96.72	5 265.66	20.21	10 194.55	1.12	4.89	0.00
2011	96.72	5 887.85	20.91	9 921.07	1.04	4.60	0.00

Source: own calculations using a stacked series of the October Household Survey (OHS) (1993 - 1999) and the General Household Survey (GHS) (2002 - 2015).

Notes: Prior Distribution = Statistics South Africa person weight for the OHS years and the design weights for the GHS years. Min. = Minimum; Max. = Maximum; | Mean | = Absolute mean; | Max. | = Absolute maximum; Share of $|\lambda|\geq 5$ = Share of constraints for the respective year with an absolute value of λ exceeding 5.

Figure 2: Total Population and Household Counts in South Africa using a Cross-Entropy Weight, 1994-2015



Source: own calculations using a stacked series of the October Household Survey (1994 - 1999) and the General Household Survey (2002 - 2015)
Notes: HHW = Household weighted; PW = Person weighted; CEW = Cross-entropy weighted; Census = Census 1996, 2001, and 2011 and Community Survey in 2007; MYPE = Mid-Year Population Estimates used for population counts outside of census years come from DataFirst back-projected population estimates for 1994 - 2001 and Statistics South Africa for 2002-2015.

9.06 million households in 1996, the CEW overestimates at 9.25 million. For the census 2007 and 2011, the CEW over- and underestimates, respectively, by 100 000 households. This is a consequence of slightly different total population counts entering into the calibration at the same time in census years. The population count enters directly into the calibration as a total count, but also indirectly via the household headship constraints. In this latter case, census year data on their headship model from SSA was used to calculate the share of each head category in the total population - and it was this indirect population count that differed slightly to the direct population count.[11] The direct population counts were sourced from official census publications and are more likely to accurately reflect weighting for non-response. In years 1996 and 2007, the direct population count exceeded the indirect population count, logically leading to an overcount on CEW's part as cross-entropy uses proportional constraints, and vice versa in 2011. Ultimately though, the calibration is reproducing its constraints. The discrepancy from the census in Table 2 is not a reflection of a poor quality

[11] In 1996, the direct population was 41.3 million and the indirect was 39.7 million. In 2007, the direct population was 48.8 million and the indirect, 48.6 million. In 2011, the direct population was 51.6 million and the indirect was 51.8 million.

calibration but of disagreement in the constraints, themselves.

As a consequence of this small dissimilitude, the HHW fits the census more accurately in the GHS period (2002 - 2015) in Figure 2 Panel (b) than the CEW; however, the CEW is able to match the census best over the entire time period. For example, the CEW accommodates the 2001 census estimate much better than the HHW. The HHW fits the census estimate in 1996 better than the CEW, but the CEW has a smoother trend over the OHS period compared to the step-wise pattern of the HHW between 1994 and 1999. Clearly the HHW has been calibrated well in census years, but the quality of the calibration is questionable in non-census years. This serves to underline the role of good quality weighting: In non-census years when population moments are not readily available, we rely upon the soundness of our weighting practise.

We can conclude that the calibration of the CEW is sound based on its resemblance to the prior **q** distribution and its strong performance in reproducing its constraints, with perhaps the exception of 1994. The CEW yields one estimate per statistic which is ideal and performs well when compared to the PW or HHW. Even if the CEW is outperformed in its fidelity to the constraints by the HHW or the PW in specific cases (e.g. census years), the CEW fits the moment information best overall for both population and household estimates.

Table 2: Total Population and Household Counts in South Africa using a Cross-Entropy Weight, 1994-2011

	Household Counts (000 000's)				**Population Counts (000 000's)**			
Year	**PW**	**HHW**	**CEW**	**Census**	**PW**	**HHW**	**CEW**	**MYPE***
1994	8.89	7.10	8.32	-	40.30	30.30	39.40	39.37
1995	8.93	9.08	8.75	-	39.70	39.70	40.40	40.35
1996	9.09	9.07	9.25	9.06	40.60	40.60	41.30	41.29
1997	9.16	9.26	9.61	-	41.40	42.20	42.10	42.14
1998	9.87	9.29	10.00	-	42.20	39.70	43.00	42.96
1999	10.80	10.80	10.50	-	43.30	43.30	43.70	43.73
2000	-	-	-	-	-	-	-	44.46
2001	-	-	-	11.21	-	-	-	45.14
2002	11.50	10.80	11.50	-	45.80	40.10	45.90	45.92
2003	12.20	11.10	11.70	-	46.40	39.50	46.50	46.46
2004	12.30	11.40	11.80	-	47.00	39.90	47.00	47.02
2005	12.60	11.80	12.20	-	47.60	41.10	47.60	47.60
2006	13.10	12.10	12.40	-	48.30	41.50	48.20	48.20
2007	13.50	12.50	12.60	12.50	48.90	42.60	48.80	48.83
2008	13.20	12.90	13.00	-	49.60	46.40	49.50	49.48
2009	13.80	13.30	13.50	-	50.20	45.80	50.20	50.15
2010	13.50	13.70	13.90	-	50.90	48.40	50.90	50.85
2011	13.50	14.20	14.40	14.50	51.60	49.30	51.60	51.57
Change 1996 - 2011	4.41	5.13	5.15	5.44	11.00	8.70	10.30	10.28

Source: own calculations using a stacked series of the October Household Survey (1993 - 1999) and the General Household Survey (2002 - 2015).

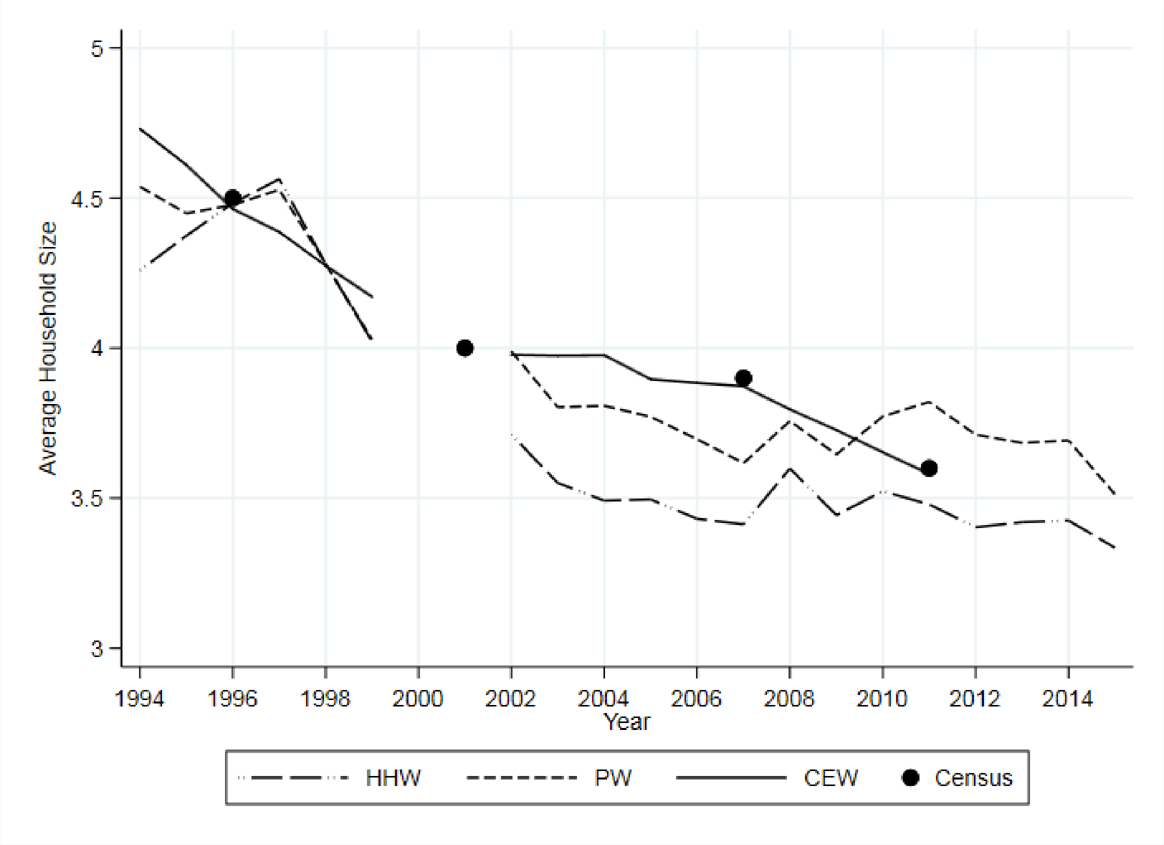
Notes: HHW = Household weighted; PW = Person weighted; CEW = Cross-entropy weighted; Census = Census 1996, 2001, and 2011, and Community Survey 2007; MYPE = Mid-Year Population Estimates; * MYPE for the period 1994 - 2001 are provided by DataFirst back projections and MYPE for the period 2002 - 2011 are provided by Statistics South Africa.

4.2 Assessing the Internal Consistency of the Cross -Entropy Weight and the Dual-Weight Scheme

Since social scientists think people and households interact in sensible meaningful ways in the real world, a desirable property of the data used to study them is that it should do the same. For our purposes then, internal consistency in household survey data means that the two main units of analysis - people and households - relate to each other in a reliable, logical way and yield estimates that do not contradict each other. This implies there is a sensible relationship between these two units in the data.

In order to investigate internal consistency, we use the statistic of average household size, following similar work by Branson and Wittenberg (2014). Figure 3 presents average household size in South Africa for the period 1994 - 2015. The CEW yields a smoother more realistic trend than the jagged, fluctuating, unstable trend estimated by both the HHW and the PW. Most importantly, the CEW fits the census estimates of household size quite precisely, unlike the SSA weights.

Figure 3: Average Household Size in South Africa using a Cross-Entropy Weight, 1994-2015



Source: own calculations using a stacked series of the October Household Survey (1993 - 1999) and the General Household Survey (2002 - 2015).
Notes: HHW = Household weighted; PW = Person weighted; CEW = Cross-entropy weighted; Census = Census 1996, 2001, and 2011 and Community Survey in 2007.

Average household size can be used to investigate internal consistency for two reasons: it is calculated using both person and household information; and, it can be calculated in two ways allowing for inspection of contradictions. Actual household size is the expected value of a derived household variable in the data set taken over the distribution of households and implied household size is determined by dividing the total population by the total household count:

$$Actual = \mathbf{E}_H[hhsize_h] \tag{8}$$

$$Implied = \frac{N}{H} \tag{9}$$

where *hhsize* is the number of people living in household h, N is the total population, H is the total household stock.

Note that in the case of the SSA weights, implied household size is calculated by dividing the *person*-weighted population by the *household*-weighted household counts. If the weighting system is internally consistent, these two methods should produce equal estimates of average household size; that is, actual should equal implied. This is evidently the case for the CEW in Figure 4. This figure reports the percentage difference between actual and implied average size. The CEW has a difference of zero for all years. By contrast, there is frequently a large discrepancy between the actual and implied average household size using the HHW and the PW. This is the case for every year in the GHS period (2002 - 2015), although the discrepancy is small in the case of the HHW in 2014 and 2015, in particular.

Based on the above section and this one, we conclude that the CEW presents a useful alternative to the existing SSA dual-weight system of the PW and the HHW. The previous section demonstrated the soundness of the cross-entropy recalibration in general and also showed that the CEW performs better than the HHW and PW in estimating population and household counts over the full post-apartheid period. This section argued that conceptually the CEW is superior since it is internally consistent in a way that the dual-weight system is not. Taken together, this means the CEW solves the two problems associated with the dual-weight system by providing singular consistent estimates and being conceptually coherent. In this way, we are able to restore respresentativness to the OHS-GHS series as the CEW allows the data to consistently represent both people and households at the same time.

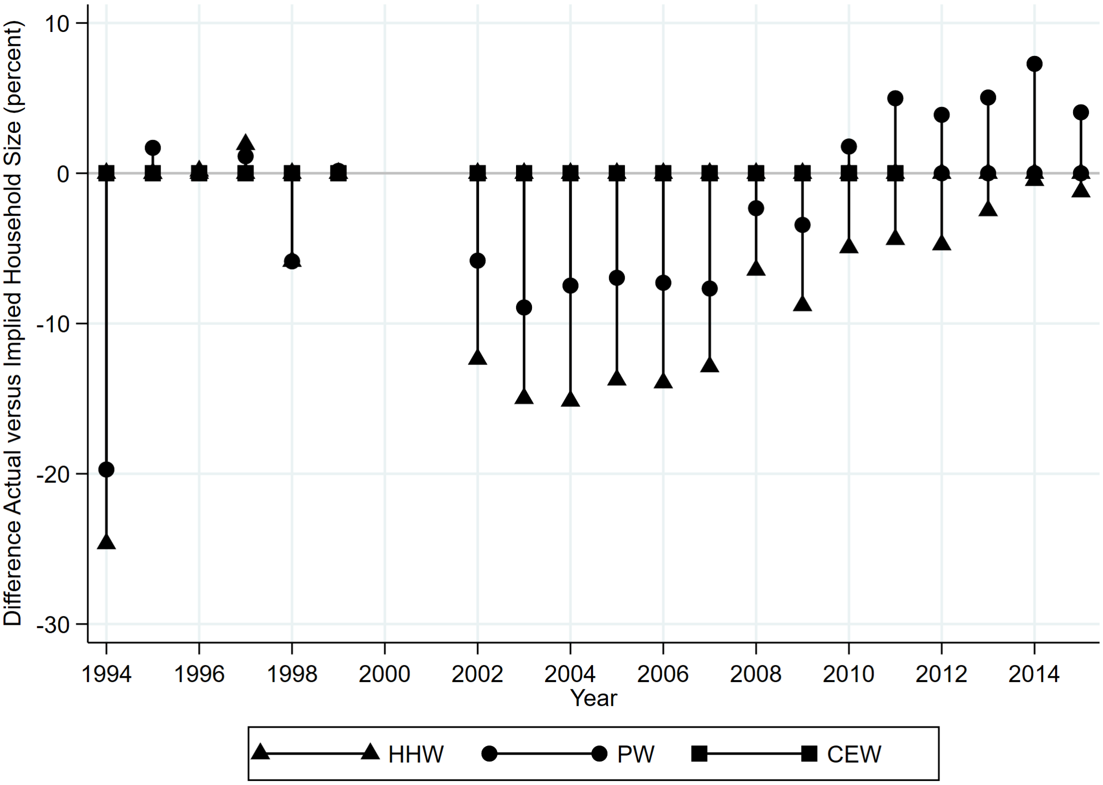
4.3 The Performance of the Cross-Entropy Weight on Non-Constraining Statistics: Single Person Households

We are confident that the CEW presents a valid alternative to the SSA dual-weight system. However, there are limits on the extent to which reweighting can improve the quality of the data. Earlier it was explained that the degree to which a sample is representative depends on factors at all phases of the data collection and dissemination process, post-sampling survey weight calibration being just one of these phases. Representativeness can also be affected by sample design, non-response, measurement error, enumerator practice, and luck surrounding the drawing of a single random sample. Weight calibration happens after all these factors have already come into play, meaning that if there are serious deficiencies with the underlying sample, reweighting is a relatively flimsy and limited tool to correct representativeness.

Limitations on the power of the reweighting mean that whilst the CEW might demonstrably perform better than the dual-weight system on certain statistics, this might not be the case for all statistics. Results so far have focused on population and household counts, statistics that explicitly entered into the calibration. In other words, population and household counts are the statistics in which we would expect these weights to perform best since they have specifically been designed to do so. A good sensitivity test of the quality of the two weighting systems then is to use statistics with which the weights have not been specifically constrained.

The share of single person households is a statistic about which researchers have long raised concerns about measurement accuracy (Branson and Wittenberg, 2014; Kerr and Wittenberg, 2015; Wittenberg and Collinson, 2008). Wittenberg and Collinson (2008) note that the proportion of single person households increases far more rapidly in national surveys than it does in the location of Agincourt, Mpumalanga, where a panel survey has been carried out for many years. Kerr and

Figure 4: Difference between Actual and Implied Average Household Size (Percent)

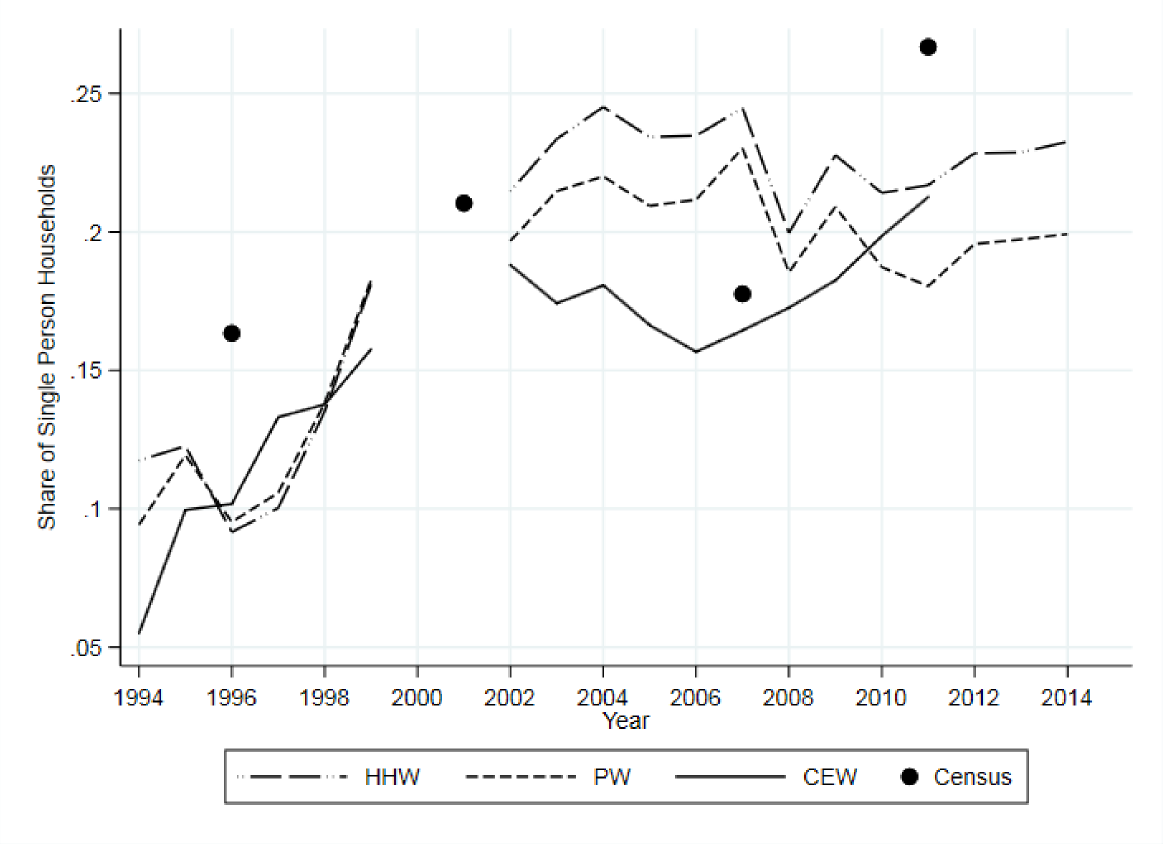


Source: own calculations using a stacked series of the October Household Survey (1993 - 1999) and the General Household Survey (2002 - 2015).
Notes: HHW = Household weighted; PW = Person weighted; CEW = Cross-entropy weighted; Actual household size calculated using the existing household size variable in the data set and weighted as specified; Implied household size for the Statistics South Africa weights is defined as the total population (person weighted) divided by the total household count (household weighted); Implied household size for the CEW is defined as total population count (CEW weighted) divided by total household count (CEW weighted).

Wittenberg (2015) argue that the share of single person households are underestimated in the OHS data set because of undersampling of small households and hostel-dwellers. These authors describe the abrupt increase in single person households between the OHS series and the Labour Force Survey series as "improbable". Such considerations suggest that there may be problems with the underlying sub-sample of single person households, undermining the power of reweighting to fix this series.

Figure 5 confirms these concerns. None of the weighted series connect with the census data points in a precise way. This is particularly prominent at the beginning and end of the period. All three of the weights underestimate the share of single person households in 1996 by more than 6 percentage points, and by 5 percentage points or more in 2011. The CEW fares better in the middle of the period, only differing from the 2007 Community Survey estimate by 1.3 percentage points. The CEW perhaps performs marginally better in that, if one were to joint the census dots, the CEW trend line would come closest to matching the shape suggested by the census points. That said, this performance is hardly convincing that the CEW is superior to either of the weights in the dual-weight system. Imprecision on the part of all three weights is likely driven by undersampling of single person households which appears to be happening throughout the OHS-GHS series.[12]

Figure 5: The Share of Single Person Households in South Africa, 1994-2015



Source: own calculations using a stacked series of the October Household Survey (1993 - 1999) and the General Household Survey (2002 - 2015).
Notes: HHW = Household weighted; PW = Person weighted; CEW = Cross-entropy weighted; Census = Census Ten Percent Samples for 1996, 2001, and 2011 and Community Survey in 2007, weighted using the Statistics South Africa household weight provided in the data set.

Nonetheless, the purpose of this figure is to make the point that reweighting is not a panacea for inadequate representativeness. The CEW will not necessarily outperform the dual-weight system on every statistic (relative to the census). Deep underlying problems with the sample - such as seriously undersampling a certain sub-sample - are not something that can be easily remedied with reweighting. What the CEW offers is not a correction to every series of statistics in the OHS-GHS series, but rather a sturdier foundational framework on which later analysis can be more reliably built. The CEW allows the series to have two dependable cornerstones, people and households, instead of one and this substantially improves the coherence of later analysis supported by this base.

[12] The problem uncovered by Kerr and Wittenberg (2015) in the OHS appears not to have been resolved even by late into the GHS series. Single person households appear in the unweighted samples of the 1996 OHS and 2011 GHS at a much lower rate than what they do in the censuses for the respective years. In the 1996 OHS, 8.89 percent of the unweighted household sample were single person compared to 16.34 percent of the census for the same year. Single person households made up 18.78 percent of the unweighted household sample in the 2011 GHS, but 26.68 percent of the 2011 census.

5 Conclusion

The strength of household surveys as knowledge resources depends on their data quality and the degree to which they can be judged as fit-for-purpose. Research conclusions from household survey data feed into policy decisions, economic and demographic planning and benchmark-setting, and a wide range of academic research, especially in the fields of consumer, labour market, and welfare studies. This means the data quality of these resources is not inconsequential and an extensive literature has been developed to control data quality at every step in the survey process, including the technique of survey weight calibration.

Calibration is an important tool for improving the degree to which a given sample is representative of its population and serves to improve the quality of inferences. However, as this paper illustrates, calibration can also create problems. This is because calibration is essentially a modelling problem and the modelling can be of better or worse quality, depending on the standard of the auxiliary information available, as well as, the modelling decisions made by the survey statistician. The South African national survey data discussed in this paper offers an insightful case study into how the data quality of an otherwise well-designed survey was compromised in the final stages by a flawed approach to calibrating the weights.

The South African case shows that the damage done by faulty calibration can be extensive: The current calibration of South African national survey data yields a weighting scheme that is not only inconsistent with its sampling practise, but also causes critical conceptual and inference problems as the main two units of analysis, people and their households, are effectively de-linked. Essentially, the data suffers from a reduction in representativeness, since it is rendered only representative of either the person or the household population at a time, depending on the weight employed. A consequence is a reduction in the number and types of research questions this data can answer. In turn, this undermines the degree to which this data can be judged as fit-for-purpose since it becomes comparatively less useful for advancing academic and policy research in South Africa.

Fortunately, the nature and timing of calibration in the survey process means that faulty calibration need not represent an incontrovertible cost to data quality. This paper describes our effort at restoring the quality of the South African survey data by recalibrating the survey weights. Cross-entropy estimation is successfully employed to recalibrate the survey weights for a stacked series of cross-sections between 1994-2011. Our alternative calibration model resolves the problems created by the existing dual weight scheme and restores the mutual representativeness of people and households to the data.

That said, it is worth stressing that reweighting is not a perfect solution for all data quality issues. There are limitations on what reweighting can solve given the quality of the raw sample, which is influenced at numerous points in the data collection process. Slightly tuning the weights will not compensate for fundamental sampling problems, like non-coverage, for example. Further, we should be precise about exactly how calibration can improve data quality in light of its strengths and weaknesses. The purpose of the reweighting is not a wholesale improvement of every data output, but rather, a strengthening of the the statistical and conceptual foundation on which later analysis is built. Throughout this paper we have argued that our cross-entropy weight achieves this. With these caveats in mind, we conclude that, overall, the cross-entropy weights improve the quality of the South African data and the degree to which it is fit for the purpose of advancing the South African research agenda

This paper illustrates with a South African case study how calibration can harm data quality, as well as, how these challenges can be overcome. Taking the time to improve the data quality of existing data resources is not only possible but also a highly worthwhile task. Doing so leverages the pay-off from resources already spent on collecting the data and benefits a broad base of stakeholders.

References

Branson, N. and Wittenberg, M. (2014), 'Reweighting South African national household survey data to create a consistent series over time: A cross-entropy estimation approach', *South African Journal of Economics* **82**(1), 19–38.
URL: *https://doi.org/10.1111/saje.12017*

DataFirst (2018), Projected Population Distribution: 1990-2001, Mimeograph, University of Cape Town, Cape Town.

Deaton, A. (1997), *The analysis of household surveys: A microeconometric approach to development policy*, The Johns Hopkins University Press, United States.

Deville, J.-C. (2000), Generalized calibration and application to weighting for non-response, *in* 'COMPSTAT', Springer, pp. 65–76.
URL: *https://doi.org/10.1007/978-3-642-57678-2*$_6$

Deville, J.-C. and Särndal, C.-E. (1992), 'Calibration estimators in survey sampling', *Journal of the American statistical Association* **87**(418), 376–382.
URL: *https://doi:10.1080/01621459.1992.10475217*

Golan, A., Judge, G. and Miller, D. (1997), The maximum entropy approach to estimation and inference, *in* 'Applying Maximum Entropy to Econometric Problems', Emerald Group Publishing Limited, pp. 3–24.

Kerr, A. and Wittenberg, M. (2015), 'Sampling methodology and fieldwork changes in the October Household Surveys and Labour Force Surveys', *Development Southern Africa* **32**(5), 603–612.
URL: *http://dx.doi.org/10.1080/0376835X.2015.1044079*

Lavallée, P. and Beaumont, J.-F. (2015), 'Why we should put some weight on weights.', *Survey Methods: Insights from the Field (SMIF)* .
URL: *https://doi.org/10.13094/SMIF-2015-00001*

Särndal, C.-E. (2010), 'The calibration approach in survey theory and practice', *Survey Methodology* **33**(2), 99–119.

Statistics South Africa (1997), October Household Survey 1997: Metadata, Technical report, Government of South Africa, Pretoria, South Africa.

Statistics South Africa (1998), October Household Survey 1998: Metadata, Technical report, Government of South Africa, Pretoria, South Africa.

Statistics South Africa (1999), October Household Survey 1999: Metadata, Technical report, Government of South Africa, Pretoria, South Africa.

Statistics South Africa (2007*a*), Community Survey 2007: Basic Results, Pamphlet, Government of South Africa, Pretoria, South Africa.

Statistics South Africa (2007*b*), Community Survey 2007: Unit Records Metadata, Technical report, Government of South Africa, Pretoria, South Africa.

Statistics South Africa (2010), Reweighting of the GHS 2002–2008 data series, Technical report, Government of South Africa, Pretoria, South Africa.

Wittenberg, M. (2008), October Household Survey 1994, Technical report, DataFirst, University of Cape Town.
URL: *https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/407/download/5258*

Wittenberg, M. (2010), ‘An introduction to maximum entropy and minimum cross-entropy estimation using stata’, *Stata Journal* **10**(3), 315.
URL: *https://doi.org/10.1177/1536867X1001000301*

Wittenberg, M. and Collinson, M. (2008), Restructuring of households in rural South Africa: Reflections on average household size in the Agincourt sub-district 1992-2003, Working Paper Number 12, Southern Africa Labour and Development Research Unit, Cape Town: SALDRU, University of Cape Town.