TRANSFORMED VARIABLES IN SURVEY SAMPLING

Raymond L. Chambers[1] and Alan H. Dorfman[2]
(1) Department of Social Statistics University of Southampton
(2) U.S. Bureau of Labor Statistics, Washington DC

KEY WORDS: balanced sampling, jackknife variance estimator, lognormal linear model, prediction, probability proportional to size sampling, residuals, simple random sampling, stratified random sampling, smearing estimator, outlier robust estimator.

Abstract
It can happen, especially in economic surveys, that we are interested in estimating the population mean or total of a variable *Y*, based on a sample, when a linear model seems appropriate, not for *Y* itself, but for a transformation (strictly monotonic function) of *Y*. In the present paper, we mainly focus on the important case where the transformation is logarithmic, but the ideas introduced here are not limited to that case. Currently available methods, based on the lognormal distribution, are reviewed, and two new methods introduced, one based on the idea of "smearing" (Duan, 1983), which do not require the lognormal assumption. Theoretical biases and variances have been calculated (although not shown herein), and suggestions made for effective sample design and for reducing sensitivity to deviant points. We evaluate and compare the different estimators we describe in an extensive empirical study on four economic populations taken from the UK Monthly Wages and Salaries Survey.

1. Introduction

Given a population of *N* units, we wish to predict the finite population total $T=\sum_{i=1}^{N} y_i$ of a variable of interest *Y*, based on a sample *s* of size *n* from that population. In addition to the sampled values of *Y*, we have auxiliary information in the form of population values x_i, i = 1, .., *N* of a covariate *X*. The standard approach to this task (see Royall, 1982) assumes a linear relationship between *Y* and *X*. Often, however, there is good reason to think that the relationship between *Y* and *X* themselves is not linear, but linear in another scale of measurement, so that we have

$$h(Y)=\beta_0+\beta_1 g(X)+\varepsilon\,, \qquad (1)$$

where β_0, β_1 are unknown parameters, we allow for a transform of *X* (possibly *X* itself), and the errors ε have mean 0 and variance σ^2. The question then becomes: how do we make an inference concerning *T*, based on the available data, using this model? Allowing for transformation of *X* does not of course by itself carry us beyond the standard linear model; the essential difficulty posed by (1) is in handling the transform of the dependent variable *Y*. In the present paper we focus mainly on the case where *h* is the (natural) logarithm *log*, and we also assume that $g(x)$ = $log(x)$, so that the special case of interest is the log-log model

$$\log(Y)=Z'\beta+\varepsilon \qquad (2)$$

where $\mathbf{Z}'=(1 \quad \log(X))$ and $\beta=(\beta_0\,\beta_1)'$.

The use of transformations in inference has a long history, and has been much studied (e.g. Deming 1984 [original publication 1943], Carroll and Ruppert 1988), but not a great deal has been done in the sampling context. Chen and Chen (1996) considered an approach based on empirical likelihood, restricting its use to attainment of confidence intervals. Their results improved on earlier coverage attained using robust variance estimators based on a linear model (Royall and Cumberland, 1985). Karlberg (2000a, 2000b) assumed the errors ε were normal (so that *Y* has a lognormal distribution) and developed predictors with negligible biases; see Section 2. Section 3 introduces two new predictors of total: a SMEARING predictor, based on ideas in Duan (1983), and a ratio-adjusted-for-sample-total (RAST) predictor. Approximations to their biases and variances have been calculated and are available in a longer version of this paper; the respective jackknife variance estimators are approximately unbiased for the variances. Vulnerability to data values that deviate from the model is noted, and modifications that improve the robustness of the proposed methods are described. Section 4 describes an extensive empirical study, evaluating several of the approaches proposed in this paper. Section 5 states conclusions.

2. Predictors based on the lognormal model

A too simple response to model (2) is to use optimal linear methods to get an (ordinary least squares) estimate $\mathbf{b}_{ols}$ of β, back-transform to get predicted values of *Y* at non-sample values, and use these to predict $T_r=\sum_r y_i$, the non-sample component of *T*. Here *r* denotes the set of non-sampled population units. This gives

$$\hat{T}_A=\sum_s y_i+\hat{T}_{r,A}=\sum_s y_i+\sum_r h^{-1}(\mathbf{z}_i'\mathbf{b}_{ols})=\sum_s y_i+\sum_r e^{\mathbf{z}_i'\mathbf{b}_{ols}}$$

the naïve back-transformation predictor of *T*.

That this is not very satisfactory is readily seen. Suppose the errors are normally distributed, $\varepsilon \sim N(0,\sigma^2)$. Then Y has a lognormal distribution, and we have $E(Y \mid X) = e^{\mathbf{Z}'\beta+\sigma^2/2}$, so that $E(T_r) = \sum_r e^{\mathbf{z}_i'\beta+\sigma^2/2}$. $\hat{T}_A$ will be biased low, since $E(\hat{T}_{r,A}) = \sum_r e^{\mathbf{z}_i'\beta+\mathbf{z}_i' \operatorname{var}(\mathbf{b}_{ols})\mathbf{z}_i/2}$, and $\mathbf{z}_i' \operatorname{var}(\mathbf{b}_{ols})\mathbf{z}_i$ is of lower order than σ^2.

Karlberg (2000a, 2000b) provides the following bias corrected estimator:

$$\hat{T}_K = \sum_s y_i + \sum_r e^{\mathbf{z}_i'\mathbf{b}_{ols}+\frac{s^2}{2}(1-a_{ii})-\frac{s^4}{4n}} = \sum_s y_i + \sum_r \hat{l}_i^{-1} e^{\mathbf{z}_i'\mathbf{b}_{ols}+s^2/2}$$

where $\hat{l}_i = e^{\frac{s^2 a_{ii}}{2}+\frac{s^4}{4n}}$, with

$$s^2 = (n-2)^{-1}\sum_s (\log(y_i) - \mathbf{z}_i'\mathbf{b}_{ols})^2,$$

$$a_{ii} = \mathbf{z}_i'(\mathbf{Z}_s'\mathbf{Z}_s)^{-1}\mathbf{z}_i, \text{ and}$$

$\mathbf{Z}_s$ denotes the matrix of sample values of Z. Under the lognormal assumption, this predictor has $O(n^{-2})$ bias, and can be expected to perform well, provided the lognormal model holds, or nearly holds.

3. The RAST and SMEARING Predictors

The preceding transformation-based predictors use bias adjustments that assume a normal distribution for the transformed variable. We introduce two new predictors that escape this restriction and have other desirable properties.

3.1 Ratio Adjustment by Sample Totals (RAST)

A method of predicting the non-sample total T_r of Y should be able to exactly recover the (known) sample total of this variable. If it does, then the method yields an unbiased predictor of this sample total, and we can anticipate that it will then also give a close to unbiased predictor of T_r, and hence of T. Let $\hat{y}_i$ denote the predicted value of y_i under the method of interest. Then this requirement translates into the condition $\sum_s y_i = \sum_s \hat{y}_i$.

None of the lognormal predictors discussed in the previous section possess this property. However, for an arbitrary estimator $\mathbf{b} = (b_0 b_1)'$ of $\boldsymbol{\beta}$, it is not difficult to modify the naïve back-transformation predictor so that it does. Put $\gamma(\mathbf{b}) = \log \sum_s y_i - \ln \sum_s e^{\mathbf{z}_i'\mathbf{b}}$ and define $\mathbf{b}^* = (b_0 + \gamma(\mathbf{b}) b_1)'$. It is then easy to see that $\sum_s e^{\mathbf{z}_i'\mathbf{b}^*} = \sum_s y_i$. The resulting predictor of T is

$$\hat{T}_{RAST} = \sum_s y_i + \sum_r e^{b_0^*+b_1^*\ln x_i} = \sum_s y_i + \frac{\sum_s y_i}{\sum_s \exp(b_0 + b_1 \log(x_i))}\sum_r \exp(b_0 + b_1 \log(x_j))$$

$$= \sum_s y_i + \frac{\sum_s y_i}{\sum_s x_i^{b_1}} \sum_r x_i^{b_1} \qquad (3)$$

which we term the Ratio Adjustment by Sample Total (RAST) predictor. More generally, we can consider using weighted sample sums in the numerator and denominator of the second term. Even more general, for the model (1), is

$$\hat{T}_{RAST} = \sum_s y_i + \frac{\sum_s w_i y_i}{\sum_s w_i h^{-1}(\mathbf{z}_i'\mathbf{b})} \sum_r h^{-1}(\mathbf{z}_i'\mathbf{b}). \quad (4)$$

We assume the weights w_i when normalized to be of order n^{-1}.

3.2 The SMEARING Predictor

For predicting a Y-value at $\mathbf{Z}$, where Y obeys the model (2), Duan (1983) suggested estimating $E(Y \mid \mathbf{Z}) = \int e^{\mathbf{Z}'\beta+\varepsilon} dF(\varepsilon)$ by

$$\hat{E}(Y \mid \mathbf{Z}) = n^{-1}\sum_s e^{\mathbf{Z}'\mathbf{b}_{ols}+R_i},$$

where the R_i are the sample residuals from the ordinary least squares (ols) fit of ln (y_i) on $\mathbf{z}_i$. For an arbitrary estimator $\mathbf{b} = (b_0 b_1)'$ of $\boldsymbol{\beta}$ this leads naturally to the corresponding SMEARING predictor of the population total:

$$\begin{aligned}\hat{T}_{SMEAR} &= \sum_s y_i + \sum_r \hat{E}(y_i \mid \mathbf{z}_i) \\ &= \sum_s y_i + \sum_r n^{-1} \sum_s e^{\mathbf{z}_i'\mathbf{b}+R_i} \\ &= \sum_s y_i + \left(\sum_r x_i^{b_1}\right)\left(n^{-1}\sum_s \frac{y_i}{x_i^{b_1}}\right). \qquad (5)\end{aligned}$$

Observe that for the log-log model, the RAST predictor in (3) is a ratio of means estimator, and the SMEARING predictor in (5), a mean of ratios estimator, in the auxiliary x^{b_1}. Again we can easily extend this to a weighted version. The generalization for the model (1) is

$$\hat{T}_{SMEAR} = \sum_s y_i + \sum_{j\in r}\sum_{i\in s} \varphi_i h^{-1}(\mathbf{z}_j'\mathbf{b} + R_i) \quad (6)$$

where the weights φ_i add to 1 and are of order n^{-1}.

We also consider the Twiced SMEARING estimator, with the predictor of non-sample total given by

$$\tilde{T}_r = \sum_r \sum_s \varphi_i h^{-1}(z_j'\hat{\beta} + R_i) + \sum_s w_{i'}\left\{Y_{i'} - \sum_s \varphi_i h^{-1}(z_{i'}'\hat{\beta} + R_i)\right\}$$

The second term amounts to an original scale residual adjustment, intended to accommodate model mis-specification.

3.3 Biases

If the working model is correct, then the RAST and SMEARING estimators have O(*1/n*) order biases. Details are available in the longer version of this paper.

Under *weighted balance*, that is, if

$$\frac{\sum_s w(x_i)^{-1} x_i^K}{n} = \frac{\sum_P x_i^K}{\sum_P w(x_i)}, \text{ for } K = 0, 1, 2, 3,\ldots$$

then the RAST and Twiced SMEARING estimators are approximately unbiased even if the working model incorrect.

To aim at balance, given a particular sample, we employed "histogram weights" w_I, where w_i is the number non-sample units j having $\left|\hat{y}_j - \hat{y}_i\right| \le R/n$, for $i \in s$, and $\hat{y}_i$ are the fitted values from a (preliminary) LS fit.

For the φ-weights, we considered two options:

(a) $\varphi_i = 1/n$, "plain vanilla"

(b) $\varphi_i = w_i$.

3.4 Dealing with outliers

All the predictors developed thus far assume that the linear model (2) for log(*Y*) in terms of log(*X*) fits well, or at least that *Y* is well behaved with respect to some underlying true model. However, the reality is that the sample data typically include a substantial number of "special" values (e.g. zero) and outliers. The logarithmic transformation effectively controls the influence of raw-scale outliers, but is then susceptible to log-scale outliers (e.g. values near zero). These values can have a large effect on back-transformed predictions.

In order to control the influence of such outliers, we use robust methods of parameter estimation. In particular, the simulation study reported on in the next section was carried out using R (Ihaka and Gentleman, 1996), and we estimated β in (2) using the *rlm* function, which is part of the *MASS* robust statistics library (Venables and Ripley, 1994). We used a biweight influence function with tuning constant c = 4.685 and calculated the standard deviation s of the residuals using the MAD estimate output by *rlm*.

For the RAST and SMEARING predictors, we can go one step further, discounting outlying terms that enter into the RAST or SMEARING adjustment terms by using the outlier robust weights $\{w_i\}$, output by *rlm*. This leads to robust versions of these predictors such as

$$\hat{T}^{rob}_{RAST} = \sum_s y_i + \sum_s w_i y_i \frac{\sum_r x_i^{b_1^{rob}}}{\sum_s w_i x_i^{b_1^{rob}}}$$

$$\hat{T}^{rob}_{SMEAR} = \sum_s y_i + \left(\sum_r x_i^{b_1^{rob}}\right)\left(\sum_s w_i \frac{y_i}{x_i^{b_1^{rob}}} \Big/ \sum_s w_i\right)$$

(where b_1^{rob} is the robust estimate of β_1 output by *rlm.*)

Thus those sample units that are effectively down-weighted as outliers in the log-scale in the course of robust estimation of the regression parameters are also down-weighted in the RAST and SMEARING adjustments. These weights are not of course the weights described in section 3.3 above to achieve weighted balance. Estimators that incorporate histogram weights will be codified with an "H", those that incorporate robust weights, with an "R". The former (and twicing, in the case of SMEAR) is meant to deal with *global* deviations from the working model; the latter is intended to handle *local* deviations from the model. It is possible to incorporate both, for example the Twiced Robustified SMEARING estimator:

SM/RH(2) =

$$\sum_s y_i + R_w \sum_r \tilde{y}_{Oi} + \sum_s w_i (y_i - R_w \tilde{y}_{Oi}),$$

where

$$\tilde{y}_{Oi} = \exp\left(\tilde{\alpha}_O + \tilde{\beta}_O \log(x_i)\right)$$

$$R_w = \frac{\sum_s w_i y_i / \tilde{y}_{Oi}}{\sum_s w_i}$$

and $\hat{\alpha}_O$ and $\hat{\beta}_O$ are outlier robust estimates with w_i is histogram weights based on the sample $\tilde{y}_{Oi}$ values.

4. Simulation Study

We carried out an extensive simulation study on four populations of businesses drawn from the UK's Monthly Wages and Salaries Survey (MWSS). These were the businesses making up two sectors of the MWSS sample, labeled A with population size N = 768, and B with N = 1005. For each sector, we considered two dependent variables Y, wages (*WAGES*) and number employed (*EMP*) at the time of the survey. For each, the dependent variable X was employment as measured on the UK Inter Departmental Business Register, the sampling frame for the MWSS, at the time of selection of the MWSS sample. This is denoted *Register EMP* below. The populations are represented graphically in Figures 1 and 2. It is readily apparent that the log-log transform yields something close to a homoscedastic linear fit, but with various anomalies peculiar to each population.

Each population was independently sampled 1000 times using (a) simple random sampling without replacement (SRSWOR), (b) size stratified random sampling (SizeSTRS), with size defined by X = *Register EMP*, (c) systematic probability proportional to size sampling (SYSPPS), with X as size variable, and finally (d) restricted "overbalanced" PPSSYS sampling that give samples that are nearly balanced with respect to inverse X weights (details in the Appendix.) In all cases sample sizes were n = 50. In the stratified case, we employed 4 strata, with strata boundaries cutting off approximately equal stratum X-totals. The "top" stratum was completely enumerated, with SRSWOR for the remaining strata. The sector B allocation was 15, 15, 15, 5, and the sector A allocation, 13,13,12,12.

For all designs, we considered 10 predictors of T. These were the Expansion Estimator (EE), the Ratio Estimator (RE), the naïve back-transform predictor (TA), the Karlberg lognormal model-based predictor (TK), the RAST predictor (RA), the SMEARING predictor (SM), and robust versions of the last four, signified by TA/R, TK/R, RA/R and SM /R respectively. In the case of stratified sampling, we used both stratified versions of these predictors (that is, within stratum estimators – "SizeSTRS/Stratified") as well as versions that ignored the strata (i.e. stratification was treated purely as a sampling device - "SizeSTRS/Unstratified"). In this latter case we also replaced EE by the across-stratum ratio estimator (RE/Across) as a more suitable comparator commensurate with "survey practice".

Additionally, for SizeSTRS/Unstratified we added on versions of the "H" estimators, namely RA/H and RA/RH, SM/H, SM/RH, SM/H(2), SM/RH(2), SM/H(2v), SM/RH(2v), where (2) refers to twicing, and "v" to the vanilla version of the choice of φ-weights (see above.) These same additional 8 estimators were also calculated for PPSSYS and overbalanced samples.

For variance estimation, we used the Jackknife for all transformation-based predictors. The conventional design-based variance estimator was used for EE, while for RE and RE/Across we used the robust variance estimator suggested by Royall and Cumberland (1981). Variances were summed by stratum for the stratified versions of the estimators. Discussion of results relating to variance and consideration of measures of performance besides root mean square error (RMSE), is contained in a longer version of this paper available from the authors.

We focus here on RMSE. Table 1 gives the "winners" for each design/population combination. It is clear that no single estimator dominates.

1. What is most notable is the impact of sample design. In all case the RMSE for SRS exceeds that of the STRS/stratified estimators, which in turn exceeds the STRS/unstratified estimators, which is about the same as the PPS estimators, which, finally, exceeds the Balanced samples.
2. The table also suggests that the use of the histogram weights and twicing is more effective in the Y = Wages populations than for Y = Employment. This makes sense since the latter provides a cleaner linear fit on the log scale. (In fact, the hypothesis of a zero quadratic term is convincingly rejected for the log-log model in the A/Y=Wages population.)

Restricting attention to just the three better designs, we give all the "near winners" in Table 2, that is, those estimators whose RMSE was within 1.1 of the smallest for the given design/population. Again, robustified estimators dominate and twicing and use of the Histogram weights have better results for the messier *WAGES* populations, and we also note:

3. Robust SMEARING (SM/R) is best for the conventional sample designs, in the sense of appearing most often (3 times in the STRS/Unstratified and 3 times in the SYPPS) but TK/R and RA/RH are close behind (both 2 and 3 times, respectively.)
4. In general, robustifying seems like a very desirable precaution.
5. The conventional expansion and ratio estimators and the naïve estimator TA are not contenders, although RE/π and TA/R each appear once.

Table 3 displays estimators in the Overbalanced samples which have RMSE within 1.1 of the smallest RMSE that occurred in the conventional designs (SYSPPS or SizeSTRS/Unstratified) for each population. The potential gain from restricting samples to those with weighted balance is readily apparent.

5. Summary

Using models for transformed data to handle non-linearity can bring gains in the prediction of finite population totals. However, outliers in the transformed scale can have a much more dramatic effect on transformation-based predictors than raw-scale outliers have on linear predictors. Our empirical results suggest that the robustified SMEARING, RAST, and Karlberg predictors are the preferred predictors for the log-log model (2), with further modification using twicing and histogram weights, where the log-log model possibly holds less strictly. In particular, it seems that SM/R, is the most consistently reliable, with TK/R and RA/RH not far behind. Also, SM/RH(2) is very effective in the messier (wage) populations. Efficiencies are strongly effected by the sample design. The RAST and SMEARING estimators can also be applied to transform models other than the logarithmic transform, and theoretical grounds have given leading us to anticipate good results. Empirical testing of their behavior, however, remains for further investigation.

References

Carroll, R. J. and Ruppert, D. (1988), *Transformation and Weighting in Regression*, New York: Chapman and Hall.

Chen, G. and Chen, J. (1996), A transformation method for fintie population sampling calibrated with empirical likelihood, *Survey Methodology*, 22, 139-146.

Deming, W. E. (1984) *Statistical Adjustment of Data*, New York: Dover [original publication 1943]

Duan, N. (1983) *Journal of the American Statistical Association*, 78, 605-610

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**, 299 – 314.

Karlberg, F. (2000a). Population total prediction under a lognormal superpopulation model, *Metron*, 53-80.

Karlberg (2000b), Survey estimation for highly skewed population in the presence of zeroes, *Journal of Official Statistics*, 16, 229-241.

Royall, R. M. (1982). Finite populations, Sampling from. Entry in the *Encyclopedia of Statistical Sciences* (eds. Johnson and Kotz), New York: Wiley.

Royall, R. M. & Cumberland, W. G. (1981). An empirical study of the ratio estimator and estimators of its variance, *Journal of the American Statistical Association* **76**, 66 - 88.

Royall, R. M. & Cumberland, W. G. (1985). Conditional coverage properties of finite population confidence intervals. *Journal of the American Statistical Association* **80**, 355 - 359.

Valliant, R., Dorfman, A.H. and Royall, R.M. (2000), *Finite Population Sampling and Inference*, New York: John Wiley

Venables, W.N. and Ripley, B.D. (1994). *Modern Applied Statistics with S-PLUS*. New York: Springer.

Table 1. Minimal Root Mean Square Error for each Design/Population

	A/Y=EMP		A/Y=WAGE		B/Y=EMP		B/Y=WAGE	
Design	**Estimator**	**minimal RMSE**	**Estimator**	**min RMSE**	**Estimator**	**minimal RMSE**	**Estimator**	**minimal RMSE**
SRSWOR	TA/R	12.30	TA/R	20.09	TA/R	11.33	TA/R	28.27
SizeSTRS Stratified	TA/R	4.19	TK/R	10.11	TK/R	6.33	RE	13.05
SizeSTRS Unstratified	SM/RH	2.18	RA/RH SM/RH(2)	7.81	SM/R	4.21	SM	10.95
SYSPPS	TK	2.87	TK	7.26	SM/R	3.98	SM/RH(2)	9.47
Over-balanced	SM	1.05	SM/RH(2)	3.01	SM/R	3.57	SM/RH(2)	5.15

Table 2. Near Best (< 1.1 Min) Root Mean Square Error for 3 Designs

	A/Y=EMP		A/Y=WAGE		B/Y=EMP		B/Y=WAGE	
Design	**Estimator**	**RMSE**	**Estimator**	**RMSE**	**Estimator**	**RMSE**	**Estimator**	**RMSE**
STRS Unstratified	SM/RH TK/R SM/R RA/RH	2.18 2.25 2.25 2.48	RA/RH SM/RH(2) SM/RH(2v) SM/RH	7.81 7.81 7.87 8.41	SM/R TK/R SM/RH	4.21 4.25 4.54	SM TK SM/R SM/RH(2v)	10.95 11.04 11.29 11.83
SYSPPS	TK SM/RH RA/R RA/RH RE/π TK/R SM/R	2.87 2.92 3.00 3.05 3.07 3.09 3.09	TK RA/RH SM/RH(2) SM SM/RH	7.26 7.70 7.81 7.86 7.95	SM/R TK/R TA/R RA/R	3.98 4.00 4.23 4.36	SM/RH(2) RA/RH SM/RH(2v) SM/R SM/H RA/H TK/R	9.47 9.53 9.90 9.97 10.17 10.19 10.33
Over-balanced	SM TK	1.05 1.09	SM/RH(2) TK	3.01 3.17	SM/R TK/R	3.57 3.67	SM/RH(2) RA/RH	5.15 5.51

Table 3. Overbalanced Samples: Estimators with RMSE < 1.1 of Minimum of RMSE in PPS, STRS/unstrat estimators

***A*/*Y*=Emp**		***A*/*Y*=Wage**	
Estimator	% of min RMSE	Estimator	% of min RMSE
SM	48	SM/RH(2)	41
TK	50	TK	44
RE/pi	59	SM	48
SM/RH(2)	66	RA/RH	51
RA/RH	71	RA	53
RA	76	SM/RH(2v)	53
SM/RH	82	SM/H(2)	58
TK/R	90	TK/R	70
SM/R	91	SM/R	71
SM/H	95	RE/pi	73
RA/R	107	SM/RH	73
		RA/H	77
		RA/R	80
		TA	103
		TA/R	110
		TA	103

***B*/*Y*=Emp**		***B*/*Y*=Wage**	
Estimator	% of min RMSE	Estimator	% of min RMSE
SM/R	90	SM/RH(2)	54
TK/R	92	RA/RH	58
RA/R	102	SM/H(2)	70
[RA/RH]	[135]	SM/RH(2)	71
[SM/RH(2)]	[139]	SM/R	72
		TK/R	83
		RA/H	85
		SM/H(2v)	91
		SM	101

Figure 1. The sector A population from the Monthly Wages and Salaries Survey

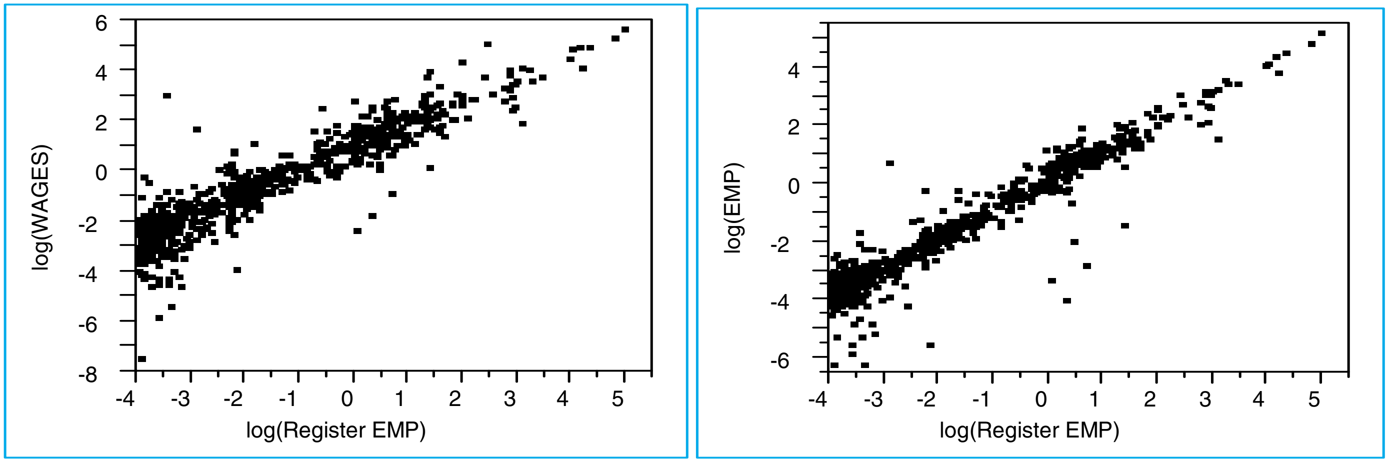

Figure 2. The sector B population from the Monthly Wages and Salaries Survey