STATISTICAL PROPERTIES OF THE APPORTIONMENT DEGREE AND ALTERNATIVE MEASURES IN BUCKING OUTCOME

Tapio Nummi*, Bikas K. Sinha** and Laura Koskela*
*Department of Mathematics, Statistics and Philosophy, Statistics Unit, University of Tampere, Finland
**Indian Statistical Institute, Kolkata, India

ABSTRACT
In the harvesting technique prevailing in Scandinavia, tree stems are converted into smaller logs immediately at harvest. Modern sawmills attempt to operate according to customers' special needs rather than only minimize the production costs. Since the annual production of saw timber in Scandinavia is in tens of millions of cubic meters, proper measuring of the goodness of the bucking outcome is of crucial importance. The outcome of the bucking operation can be considered as a multidimensional table of tree species, quality grades, prices and length and diameters classes. The prevailing method to measure the outcome is the so-called apportionment degree, which is calculated from the relative portions of the observed and target tables. However, this measure has severe drawbacks. E.g. it gives the same weight for each log class. Therefore, for example, the effect of the shape of the distributions is completely ignored. In this study we present some basic results of the statistical properties of the apportionment degree and present some alternative means to measure the bucking outcome. Also a simulation study is carried out to illustrate the relative performance of the measures presented.

Key words: Apportionment degree, forest harvesting, frequency Chi-Square, simulation.

MSC: 62P12

RESUMEN
El propósito de este trabajo es comenzar el análisis estadístico del índice del prorrateo, esta es una medida usada en el contexto de la tala en bosques para evaluar el ajuste entre la demanda y distribución del suministro de la leña. Ha habido algunos esfuerzos por entender este índice, pero una base teórica seria todavía falta. Nosotros discutimos brevemente la literatura existente y procedemos a investigar las propiedades del índice desde un punto de vista distribucional. Este es fundamentalmente un artículo exploratorio y nosotros sólo enfocamos los casos de dos y tres clases de leña, es decir, locaciones. En el caso de dos clases usamos la distribución beta para las variables aleatorias relativas a la salida (output) del rendimiento; en tres locaciones al azar se asume que los rendimientos relativos siguen la distribución de Dirichlet singular. Usando esta formulación es posible entender las propiedades estadísticas del índice del prorrateo.

1. INTRODUCTION

The general objective in harvesting is to maximize the value of the timber obtained for further processing. Optimization of harvesting requires that several phases in a production chain are successfully combined. In the harvesting technique prevailing, in Scandinavia, tree stems are converted into smaller logs immediately at harvest. High-class measuring and computing equipment have been developed, making possible computer-based optimization of crosscutting in harvesters. In modern harvesters tree stems are run in sequence through the measuring equipment and simultaneously the harvester's computer receives the length and diameter data from sensors. If the whole stem is measured before crosscutting we may apply the techniques discussed e.g. in Näsberg (1985) to find the optimal cutting patterns on the stem. However, in practice the first cutting decisions have to be made under incomplete stem information and we must compensate the unknown part of the stem by predictions (see e.g Liski and Nummi (1995)).

An admissible cutting pattern is a set of cutting points $0 = x_1 < x_2 < \ldots < x_R$ such that the length of the rth log

$$l_r = x_r - x_{r-1} \in [l_{min}, l_{max}] \text{ and } d(x_r) \geq d_T > 0$$

for r = 2, 3, R, where $x_1 = 0$ is at the butt of a tree, l_{min} is the minimum and l_{max} the maximum length of a log and d_T is the minimum acceptable log diameter. Marking for bucking is the problem of converting a single tree stem into logs in such a way that the total stem value (price, volume etc.) for logs is maximized (see Näsberg 1985, Chapter 3).

We can classify a log with a small end diameter d(x) and length l (index r dropped) to one of the $m \times n$ classes according to the following classification

$$d_i \leq d(x) < d_{i+1} \quad \text{and} \quad l_j \leq l < l_{j+1},$$

where d_i, i = 1,…,n and l_j, j = 1,…,m are given diameter and length limits. Then we may for example specify the price of each diameter and length combination d_i, l_j of logs. Denote these as the $m \times n$ price matrix **P**, where the element p_{ij} of **P** is the price of the log at log class d_i, l_j. However, it is well known that optimization of price only may yield very undesiderable log distributions from the sawmills point of view. Nowadays sawmills aim to operate more on customers special needs rather than maximizing price or minimizing the production costs only. In fact we may have simultaneously many targets. Especially we may have a matrix of frequencies jointly with a matrix of prices. We may define the target amount of logs for each diameter and length combinations. Denote these as the $m \times n$ target matrix **T**. Similarily we can classify the outcome of the actual bucking operation to elements of the $m \times n$ frequency matrix **O**. Then the measures studied here are simple functions of the actual output **O**, the target **T** and the log prices **P**.

2. MEASURING THE BUCKING OUTCOME

2.1. The Apportionment degree

The so-called apportionment degree widely used in harvesting is defined for a fixed quality class as follows

$$A = \left(1 - 0.5 \times \sum_{i=1}^{n}\sum_{j=1}^{m}\left| o_{ij}^{*} - t_{ij}^{*} \right| \right),$$

where $o_{ij}^{*} = \dfrac{o_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m} o_{ij}}$ and $t_{ij}^{*} = \dfrac{t_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m} t_{ij}}$, where o_{ij} and t_{ij} are elements of the outcome and target matrices, respectively. With this measure we can compare the relative proportions of the output and target tables. The apportionment degree A gives a value between 0 to 1, where the value A = 1 corresponds to perfect match of the tables. After some simple manipulations we can show that A can be rewritten as

$$A = \sum_{i=1}^{n}\sum_{j=1}^{m} \delta_{ij}, \tag{1}$$

where $\delta_{ij} = \min(o_{ij}^*, t_{ij}^*)$. This measure was first introduced by the swedish mathematician Bergstrand in the mid 1980s, when first steps in developing automatic bucking systems were taken. However very little is known of the statistical properties of the apportionment degree A.

It is easy to give a price-weighted version of A. Then we compute

$$A_p = \sum_{i=1}^{n}\sum_{j=1}^{m} \delta_{ij} p_{ij}^{*},$$

where $p_{ij}^{*} = \dfrac{1}{\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}}$ is the relative price of the log at given log class. However, A_p is no longer in the same magnitude as A. We first note that

$$\rho_{\delta,p^*} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} \delta_{ij} p_{ij}^{*} - \dfrac{A}{nm}}{\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{m} \delta_{ij}^{2} - \dfrac{A^2}{nm}\right)\left(\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{*2} - \dfrac{1}{nm}\right)\right]^{0.5}} \leq 1$$

and, hence,

$$A_p = \sum_{i=1}^{n}\sum_{j=1}^{m}\delta_{ij}p_{ij}^{*} \le \frac{A}{nm} + \sqrt{\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\delta_{ij}^{2} - \frac{A^2}{nm}\right)\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\delta_{ij}^{*2} - \frac{1}{nm}\right)}\,.$$

We denote $\sigma_\delta = \sqrt{\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\delta_{ij}^{2} - \frac{A^2}{nm}\right)}$ and $\sigma_{p^*} = \sqrt{\left(\sum_{i=1}^{n}\sum_{j=1}^{m}p_{ij}^{*2} - \frac{1}{nm}\right)}$. Then it is easy to see that the values of A_p are at the interval

$$A_p \in \left[\, 0, \frac{A}{nm} + \sigma_\delta\sigma_{p^*} \right].$$

Then instead of A_p we can calculate

$$A_s = \frac{nmA_p}{A + nm\sigma_\delta\sigma_{p^*}},$$

which is of the same magnitude as the original A. Note that our choise $\rho_{\delta p^*} = 1$ generally overestimates the true value of ρ_{δ,p^*}. This may imply that on the average the scaled statistic A_s underestimates the true apportionment degree.

It is now easy to make some observations concerning A_s. First if δ and p* are independent we note that $A_s = nmA_p/A$. Similarly if $\sigma_\delta \approx 0$ or $\sigma_{p^*} \approx 0$ we note that in both cases $A_s = nmA_p/A$. These correspond to situations where price is approximatelly uniform or the disparity between demand and supply is more or less uniform, respectively.

2.2. Analysis with standard statistical measures

One of the most common measures to test the fit between two distributions is the χ^2-test. By using our notations this statistic is defined as

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}(o_{ij} - t_{ij})^2 / t_{ij},$$

which under certain conditions follows the χ^2-distribution with nm - 1 degrees of freedom. Note that here we use nm instead of nm - 1 degrees of freedom as an approximation since nm is in practical situations appropriately large. A price-weighted version of the statistic can be written as

$$\chi^2(p^*) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}p_{ij}(o_{ij} - t_{ij})^2 / t_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m}p_{ij}} \qquad (2)$$

It can be shown see e.g. Rao (1973) that the distributions of this statistic can be approximated by weighted χ^2-distribution.

2.2.1. Relation to the apportionment degree in the unweighted case

We first write A as in (1), and

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(o_{ij}^2 / t_{ij} - \frac{N}{nm} \right), \qquad (3)$$

where we assume $N = \sum_{i=1}^{n}\sum_{j=1}^{m}t_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m}o_{ij}$. Then define

$$o_{ij} = t_{ij} + \varepsilon_{ij}, \ \forall i,j$$

where we assume $\sum_{i=1}^{n}\sum_{j=1}^{m} \in_{ij} = 0$,

$$\sum^{I} = \sum_{\{(i,j)|o_{ij}<t_{ij}\}} \text{ and } \sum^{II} = \sum_{\{(i,j)|o_{ij}\geq t_{ij}\}}$$

By using these notations we write the χ^2-statistic in (3) as

$$\chi^2 = \sum^{I}\left[\frac{(t_{ij}+\in_{ij})^2}{t_{ij}} - \frac{N}{nm}\right] + \sum^{II}\left[\frac{(t_{ij}+\in_{ij})^2}{t_{ij}} - \frac{N}{nm}\right]$$

$$= \sum^{I}\left[(t_{ij}+2\in_{ij}) + (\in_{ij}^2/t_{ij}) - \frac{N}{nm}\right] + \sum^{II}\left[(t_{ij}+2\in_{ij}) + (\in_{ij}^2/t_{ij}) - \frac{N}{nm}\right].$$

Now we can write

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} t_{ij} + \sum^{I}\left(\in_{ij}^2/t_{ij} - \frac{N}{nm}\right) + \sum^{II}\left(\in_{ij}^2/t_{ij} - \frac{N}{nm}\right).$$

This can be further written as

$$\chi^2 = N(A-1) - \sum^{I}\in_{ij}^2 + \sum_{i=1}^{n}\sum_{j=1}^{m}\in_{ij}^2/t_{ij},$$

where we note that in our notations

$$A = \frac{1}{N}\left(\sum_{i=1}^{n}\sum_{j=1}^{m} t_{ij} + \sum^{I}\in_{ij}\right).$$

2.2.2. Distribution of the weighted χ^2-statistic

The price-weighted χ^2-statistic in (2) can be written as a sum

$$\chi^2(p^*) = \sum_{i=1}^{n}\sum_{j=1}^{m} u_{ij}^2,$$

where

$$u_{ij} = \sqrt{p_{ij}^*}\,\frac{o_{ij}-t_{ij}}{\sqrt{t_{ij}}}.$$

It is easy to see that

$$u_{ij} \sim N(0,\ p_{ij}^*)$$

and

$$\frac{1}{\sqrt{p_{ij}^*}}u_{ij} = z_{ij} \sim N(0, 1).$$

Now $\chi^2(p^*) = \sum_{i=1}^{n}\sum_{j=1}^{m} u_{ij}^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} z_{ij}^2 p_{ij}^*$ and hence the distribution of $\chi^2(p^*)$ is a weighted sum of independent χ_1^2-variables. The distribution of $\chi^2(p^*)$ can be approximated by

$$\chi^2(p^*) \approx a\chi^2_b. \quad (4)$$

Now a and b can be solved from the first two moments of $\chi^2(p^*)$. The expected value is

$$E(\chi^2(p^*)) = 1.$$

For the variance we first note that the weights lie between the values

$$p^*_{min} \le p^*_{ij} \le p^*_{max}, \ \forall i,j.$$

Then we note that

$$p^*_{min}\chi^2_{nm} < \chi^2(p^*) < p^*_{max}\chi^2_{nm},$$

and therefore the variance is bounded by

$$2nmp^{*2}_{min} < V(\chi^2(p^*)) < 2nmp^{*2}_{max}.$$

Then one approximation would be

$$V(\chi^2(p^*)) \approx 2nm(p^{*2}_{min} + p^{*2}_{max})/2 = nm(p^{*2}_{min} + p^{*2}_{max}).$$

We can now solve for a and b by equating the mean and variance of both sides in (4):

$$ab = 1$$

and

$$2a^2b = nm(p^{*2}_{min} + p^{*2}_{max}) = T, \text{ say.}$$

It follows that

$$\hat{a} = T/2$$

and

$$\hat{b} = 2/t,$$

where $\hat{a}$ and $\hat{b}$ are estimates of multiplier and degrees of freedom of the approximation (4), respectively.

2.2.3. A computational example

Assume that the target matrix is given in the Table 1. Then, for example, the target of the length of 430 cm and the top diameter of 160 mm logs is 28 objects. Assume that the actual output matrix is given in the Table 2. In fact the output matrix is obtained from the target matrix by randomly dropping 15 percent of logs from the target table.

Table 1. Target matrix.

Top diam (mm)	length (cm)					Total
	430	460	490	520	550	
160	28	16	58	45	45	192
200	37	17	65	45	37	201
240	17	49	37	44	55	202
280	22	39	39	44	59	203
340	19	30	47	54	52	202
Total	123	151	246	232	248	1000

Table 2. Output matrix.

Top diam (mm)	length (cm)					Total
	430	460	490	520	550	
160	23	12	52	39	41	167
200	33	11	61	39	27	171
240	12	39	30	38	49	168
280	14	34	33	36	56	173
340	8	30	42	47	44	171
Total	90	126	218	199	217	850

Table 3. Price matrix 1.

Top diam (mm)	length (cm)					Total
	430	460	490	520	550	
160	100	103	105	108	109	525
200	124	128	130	134	135	651
240	144	148	151	156	157	756
280	156	161	164	168	170	819
340	160	165	168	173	174	840
Total	684	705	718	739	745	3591

Next we investigate the fit between these two tables. The apportionment degree of these tables is A = 0.963. Since the value of the statistic is very close to 1 the fit between the two matrices is very good. The ordinary χ^2-test statistic gives the associated p-value 0.176. This comparison also shows that the fit between the target and the observed matrices is very good.

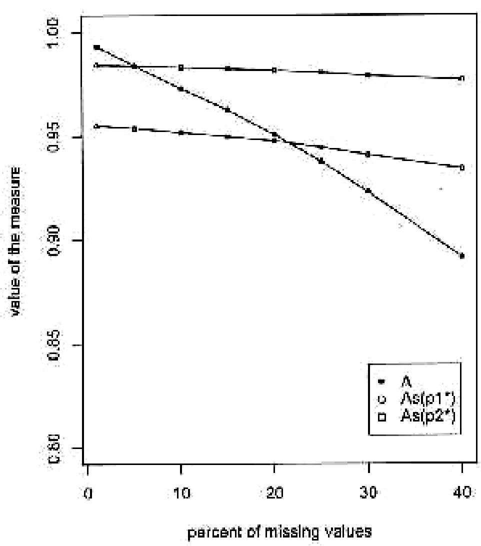
We may also specify the price of each diameter and length combination of logs. Here we use two price tables denoted by P_1 and P_2. The price matrix 1 is given in the Table 3 and P_2 is simply the matrix transpose of P_1. The scaled price-weighted apportionment degrees are $A_s(p_1^*)$ = 0.947 and $A_s(p_2^*)$ = 0.982, respectively. It is easy to see that also in this case the fit is very good, however a slightly better fit is obtained when the price matrix P_2 is used.

The price-weighted versions $\chi^2(p_1^*)$ and $\chi^2(p_2^*)$ gave the associated p-values 0.152 and 0.242, respectively. This also indicates a slightly better fit attained when using the price matrix P_2.

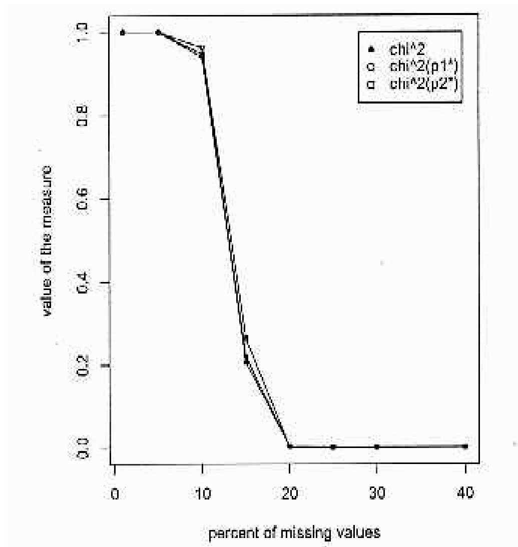
3. A SIMULATION STUDY

In this section we conduct a simulation study to investigate the performance of the apportionment degree, χ^2-test statistic and their price-weighted versions to measure the fit between target and output matrices. We take the target matrix in the Table 1 as a starting point. Next we randomly deleted 1%, 5%, 10%, 15%, 20%, 25% 30% and 40% in turn of the logs in the target table and this experiment is repeated 100 times at each percentage point. At each point the values of the apportionment degree, the χ^2-statistic and their price-weighted versions were calculated. For χ^2-statistics also the associated p-values were calculated.

Then mean curves of A, $A_s(p_1^*)$ and $A_s(p_2^*)$ are given in Figure 1a and mean curves of the p-values of χ^2, $\chi^2(p_1^*)$ and $\chi^2(p_2^*)$ are given in Figure 1b. From Figure 1a we observe that the average performance of the Apportionment degree A and its price-weighted versions $A_s(p_1^*)$ and $A_s(p_2^*)$ are approximately linear as a function of randomly generated missing values. The average decrease is greatest for the Apportionment degree A. For price-weighted versions $A_s(p_1^*)$ and $A_s(p_2^*)$ the decrease is approximatelly the same, but the values computed for $A_s(p_2^*)$ are at somewhat higher level. It is remarkable that although the percentage of generated missing values is relatively high, the Apportionment degree indicates quite good fit. For example, if the percentage of generated missing values is as high as 40 % the average value of the Apportionment index is approximatelly 0.89 with very narrow range of values (see Figure 2 in the Appendix). Thus even large departures from the target table gave quite high values of the measure. This is not a very good property of a statistic, but it may make sense in practical applications where the values of the target table may not be possible to attain exactly. However, some kind of rough measure is needed to relate the target table to the observed one. Note that this measure compares only the relative values of the observed and target tables. Therefore large departures in the absolute values may not be noticed.



(a) Mean curves of A, $A_s(p_1^*)$ and $A_s(p_2^*)$.



(b) Mean curves of p-values of χ^2, $\chi^2(p_1^*)$ and $\chi^2(p_2^*)$

Figure 1. Mean curves of the simulation study.

Statistically the χ^2-statistic and its price-weighted versions performed better. When the percentage of missing values is 20 % or more, these statistics clearly reject the null hypothesis of the fit of the observed and target tables (Figure 1b). The performance of each of the χ^2-statistics follows approximately the similar pattern (see also Figure 3 in the Appendix).

The simulation in this section was carried out by using R computing environment (see e.g. http://www.r-project.org/).

4. CONCLUDING REMARKS

In this paper we study the use of the apportionment degree and the χ^2-statistic and their price-weighted versions when measuring the fit of the output and target tables. This comparison shows that the apportionment degree clearly measures the difference in relative values whereas the χ^2-statistic also observes the differences in absolute values. The idea of using prices as weights leads us to the use of the theory of index numbers for measuring the goodness of the bucking outcome which is a topic for future research.

5. ACKNOWLEDGMENTS

This work was financed by the Academy of Finland under the project entitled "Advanced Methods for Computer-Aided Bucking of Scots Pine". We are thankful to an anonymous referee for critically reading the manuscript and offering helpful suggestions which improved the presentation of our ideas.

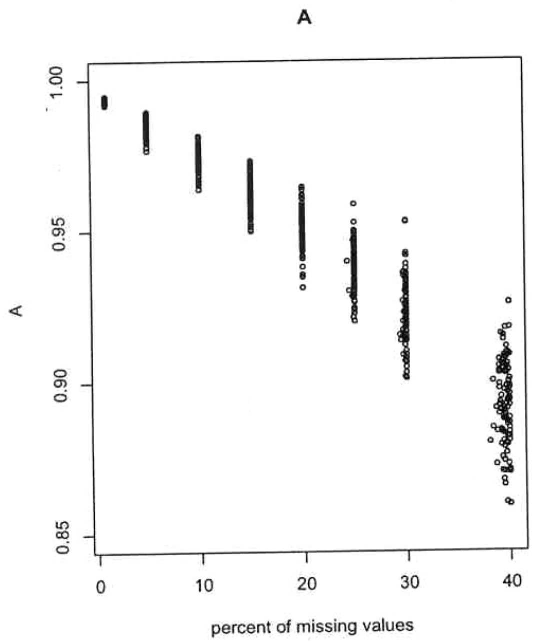
REFERENCES

LISKI, E.P. and T. NUMMI (1995): “Prediction of tree stems to improve efficiency in automatized harvesting of forests”, **Scandinavian Journal of Statistics**, 22, 255-269.

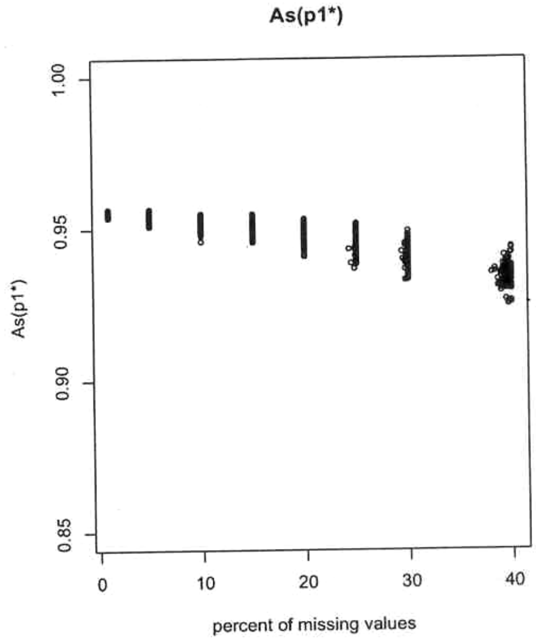
NÄSBERG, M. (1985): “Mathematical programming models for optimal log bucking”, **Linköping Studies in Science and Technology**, Dissertation No. 132.

RAO, C.R. (1973): **Linear Statistical Inference and its Applications**, Second Edition, Wiley, New York.

APPENDIX



(a) Results of A.



(b) Results of $A_s(p_1^*)$.

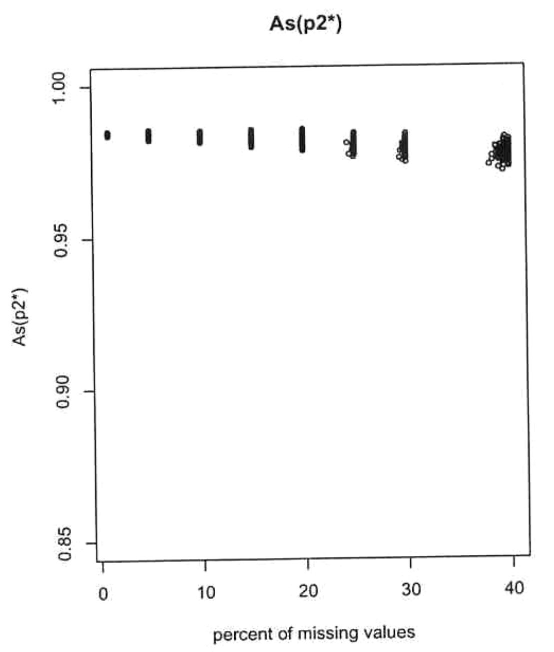
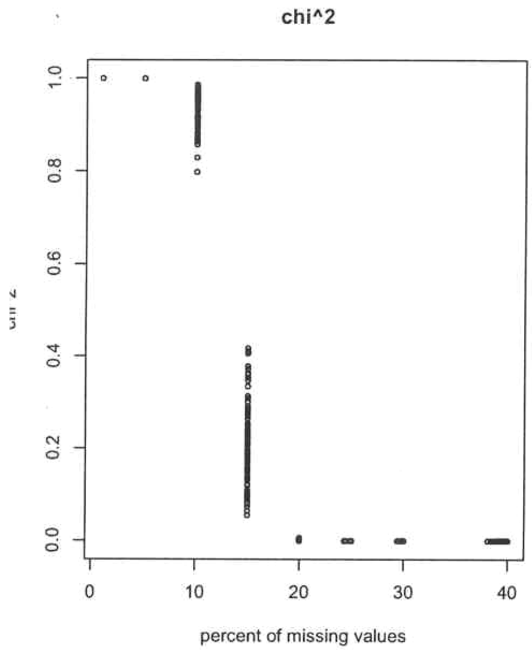
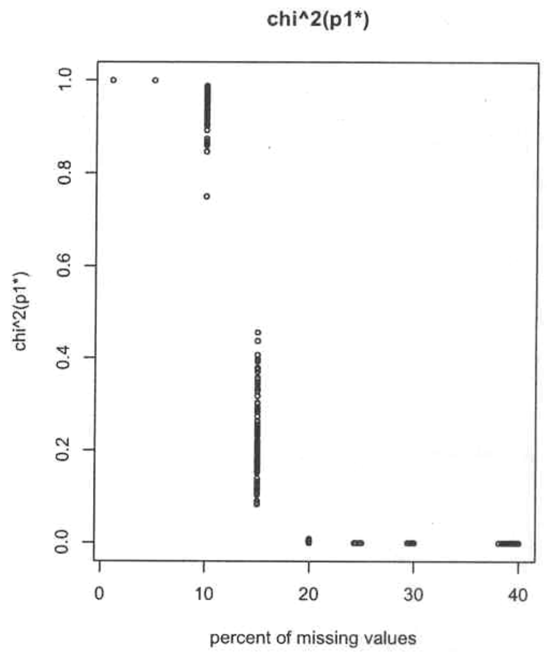


Figure 2. Results of the simulation study for Apportionment degree and its scaled price-weighted versions.



(a) Associated p-values of χ^2.



(b) Associated p-values of $\chi^2(p_1^*)$.

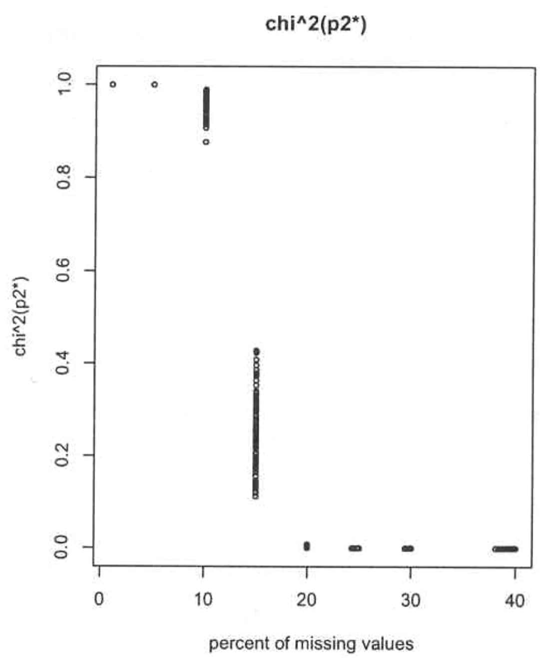


Figure 3. Results of the simulation study of ordinary χ^2-test statistic and its price-weighted versions.