LUND UNIVERSITY

Are the Sweden Democrats really Sweden's largest party?

A maximum likelihood ratio test on the simplex

Bergman, Jakob; Holmquist, Björn

Are the Sweden Democrats really Sweden's largest party?
A maximum likelihood ratio test on the simplex

J. Bergman*, B. Holmquist

Department of Statistics, Lund University, Box 743, SE-220 07 Lund, Sweden

Abstract

We introduce a maximum likelihood ratio test to test if a specific proportion is the greatest in a multinomial situation with a single measurement. The test is based on partitioning the parameter space and utilising logratio transformations.

Keywords: Isometric logratio transformation, Largest party, Maximum likelihood ratio test, Political polls, Polls, Sweden, Simplex
2010 MSC: 62F03, 62H15, 62P25

1. Introduction

On 20 August 2015, the Swedish newspaper *Metro* ran the headline 'Now the Sweden Democrats are Sweden's largest party' (our translation) across its front page (Wallroth, 2015). From a journalistic point of view the headline was not surprising: 10 years ago the nationalistic party the Sweden Democrats (SD) had a voter share of 1–2% and was hardly ever reported in the polls, and now there was a poll that gave the party the largest voter share of any party. A remarkable change indeed. However, from a statistical point of view the headline was intriguing: how do we test such a claim? Any introductory text book in statistics will tell you how to test if a proportion is greater than a specified value in a binomial situation. But in this case there is no specified value to test, and furthermore, Sweden has a multiparty system with, in practice, 8–10 competing parties, so this is a multinomial situation. So, how can we test if a specific share is greater than all the others?

One immediate approach would be to perform pairwise tests of the specific share against each of the others. However, to attain an overall level of significance, these tests need to be adjusted, e.g. with a Bonferroni correction. Apart from the general lack of elegance of such an approach, the procedure becomes less attractive when the number of parties increases; in a ten-party situation nine tests would be needed and for each test the significance level would have to be a mere 0.0055 for the overall significance level to be 0.05. A more serious objection is that such procedures do not incorporate the implicit structure of the observed shares or frequencies; due to the fact that the shares need to sum to 1 or the frequencies to n, respectively, they are not independent but negatively correlated.

Instead of multiple tests, we would like one single test. We propose a maximum likelihood ratio test utilising the inherent properties of shares (proportions) to test the hypothesis. In Section 2 we introduce some notation, formalise the problem and discuss the properties of the parameter space, in Section 3 we derive the test and its properties. We apply the test to the newspaper article above in Section 4.

---

*Corresponding author
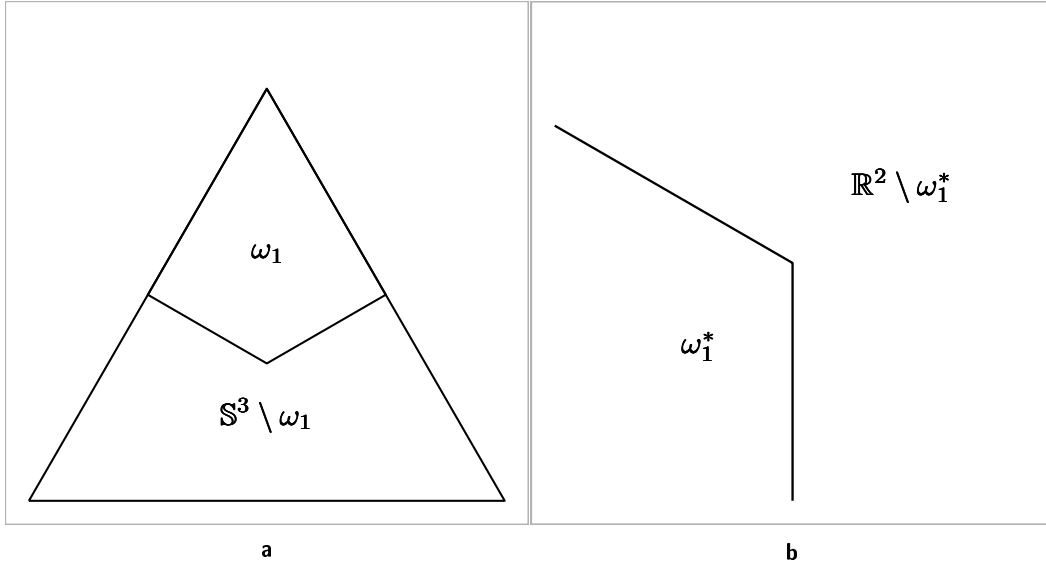*Email addresses:* `jakob.bergman@stat.lu.se` (J. Bergman), `bjorn.holmquist@stat.lu.se` (B. Holmquist)

Figure 1: The parameter space $\mathbb{S}^3$ is shown in (a) partitioned into the subspace ω_1, where p_1 is the largest part, and $\mathbb{S}^3 \setminus \omega_1$, where p_1 is not the largest part. The top vertex corresponds to $\boldsymbol{p} = (1, 0, 0)'$, the bottom left to $\boldsymbol{p} = (0, 1, 0)'$, and the bottom right to $\boldsymbol{p} = (0, 0, 1)'$. The boundary between ω_1 and $\mathbb{S}^3 \setminus \omega_1$, is the line from $\boldsymbol{p} = (1/2, 1/2, 0)'$, via $\boldsymbol{p} = (1/3, 1/3, 1/3)'$ to $\boldsymbol{p} = (1/2, 0, 1/2)'$. In (b) the corresponding parameter space in $\mathbb{R}^2$, partitioned into $\omega_1^* = \mathrm{ILR}(\omega_1)$ and $\mathbb{R}^2 \setminus \omega_1^*$, is shown.

2. Voter shares and the simplex

Let $\boldsymbol{p} = [p_j]$ denote the vector of voter shares of the D parties in the electorate. (If there is a multitude of very small parties, the Dth share can represent the sum of all small parties.) Since $\boldsymbol{p}$ is non-negative and must sum to 1, the parameter space of $\boldsymbol{p}$ is the D-part simplex $\mathbb{S}^D$. Given a simple random sample of n respondents, the number of voters for each party $\boldsymbol{X}$ is a multinomial random variable with parameter $\boldsymbol{p}$. (Of course, $\boldsymbol{X}$ actually has a multivariate hypergeometric distribution, but we will assume that the population is large enough for the multinomial distribution to be an acceptable approximation.)

The statement that the ith share p_i is the greatest of the D shares is a relative statement, which, however, has absolute implications: a necessary condition is that $p_i > 1/D$ and a sufficient condition is that $p_i > 1/2$ (see Appendix A for proofs). We believe though that it is easier to consider the entire parameter space than to try to find explicit expressions for p_i. This means testing the hypotheses

$$
\begin{aligned}
H_0 &: \boldsymbol{p} \in \mathbb{S}^D \setminus \omega_i \\
H_1 &: \boldsymbol{p} \in \omega_i
\end{aligned}
\tag{1}
$$

where ω_i is the subspace of $\mathbb{S}^D$ in which the ith part (share) is the greatest. The boundary between the two subspaces is the line, plane etc. where $p_i = p_j$ for at least one $j \neq i$ and all other parts are smaller. As an illustration, the parameter space $\mathbb{S}^3$ is depicted in Fig. 1(a) as a ternary diagram.

However, the simplex can pose practical problems due to the constraints on the parameters. Aitchison (1982) introduced the logratio transformations to resolve some of these issues. One popular choice of such transformation is the isometric logratio (ILR) transformation (Egozcue et al., 2003). It resolves the summation constraint of the simplex and transforms the problem to the real space $\mathbb{R}^{D-1}$. As an illustration, the subspaces in $\mathbb{R}^2$ corresponding to $\mathbb{S}^3 \setminus \omega_1$ and ω_1 are depicted in Fig. 1(b). There are many different conceivable ILR transformations, one example is the vector $\boldsymbol{y} = [y_j]$ where

$$
y_j = \frac{1}{\sqrt{j(j+1)}} \log \frac{\prod_{k=1}^{j} p_k}{p_{j+1}^j}, \qquad j = 1, \ldots, D-1.
\tag{2}
$$

3. A maximum likelihood ratio test

We propose that (1) is tested using a maximum likelihood ratio (MLR) test. (Here we follow the terminology used by e.g Garthwaite et al. (2002, Sec. 4.6).) This means finding the maximum value of the likelihood in the restricted parameter space under H_0 and comparing this with the maximum value if the parameter space is not restricted. As the sample consists of only one observation, the likelihood function equals the probability function

$$L(\boldsymbol{p}|\boldsymbol{x}) = \frac{n!}{x_1! \cdots x_D!} p_1^{x_1} \cdots p_D^{x_D}. \tag{3}$$

The maximum in the unrestricted space is simply the value of the likelihood of the ML estimate $\hat{\boldsymbol{p}} = \boldsymbol{x}/n$. In the restricted parameter space under H_0, the ML estimate is $\boldsymbol{p}^* = \arg\max_{\{\boldsymbol{p} \in \mathbb{S}^D \setminus \omega_i\}} L(\boldsymbol{p}|\boldsymbol{x})$. Assuming that x_i, the part corresponding to p_i, is the largest in the observed vector $\boldsymbol{x}$, the restricted estimate $\boldsymbol{p}^*$ will be a point on the boundary of $\mathbb{S}^D \setminus \omega_i$. This means maximising (3) over $\boldsymbol{p}$ subject to

  (a) $p_i \leq p_j$, for some $j \neq i$,
  (b) $p_j > 0$, for $j = 1, \ldots, D$, and
  (c) $p_1 + \cdots + p_D = 1$.

Normally, $\boldsymbol{p}^*$ will have to be estimated numerically. The optimisation is simplified if the last two constraints are removed by transforming the problem to the real space using an ILR transformation. The first constraint (a) can then be accomplished by maximising (3) over the complement set to the set determined by the linear inequality $\boldsymbol{u}_i\boldsymbol{y} \geq \boldsymbol{0}$, i.e. $\bigcup_{j \neq i}\{\boldsymbol{u}_j\boldsymbol{y} \geq \boldsymbol{0}\}$. The matrices $\boldsymbol{u}_j$ will depend upon the exact choice of ILR-transform. However, one may utilise the fact that if $\boldsymbol{p} \in \omega_1$ then $\log p_1/p_2 > 0$, $\log p_1/p_3 > 0$ etc. in solving these inequalities. If (2) is used, as suggested by Egozcue et al. (2003, p. 296), and $D = 5$, then e.g.

$$\boldsymbol{u}_1 = \begin{bmatrix} \sqrt{2/1} & 0 & 0 & 0 \\ 1/\sqrt{2} & \sqrt{3/2} & 0 & 0 \\ 1/\sqrt{2} & 1/\sqrt{6} & \sqrt{4/3} & 0 \\ 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{12} & \sqrt{5/4} \\ 3/\sqrt{2} & 5/\sqrt{6} & 5/\sqrt{12} & 5/\sqrt{20} \end{bmatrix}.$$

The test statistic is

$$\lambda = -2\big(\log L(\hat{\boldsymbol{p}}) - \log L(\boldsymbol{p}^*)\big) = -2\sum_{j=1}^{D} x_j \log \frac{x_j}{np_j^*}. \tag{4}$$

If H_0 is true and $\boldsymbol{p}$ is on the boundary of $\mathbb{S}^D \setminus \omega_i$, then (4) will be 0 with probability δ equal to the proportion of the probability mass located in $\mathbb{S}^D \setminus \omega_i$. The exact value of δ depends on $\boldsymbol{p}$, but it will typically be close to $1/2$, unless $\boldsymbol{p}$ is close to a point where $d \geq 3$ parts are equal, in which case δ will be close to $(d-1)/d$. Conditionally on positive values, hence with probability $1 - \delta$, the statistic in (4) has a distribution related to the asymptotic chi-square distribution for log likelihood ratio statistics. This asymptotic distribution will be depending on d, and we found, by simulations, $d\lambda/2$ to have a χ^2-distribution with one degree of freedom, to be a fairly accurate approximation of the distribution. For $d = 2$ this would be exact for the asymtotics of likelihood ratio statistics, where the one degree of freedom follows from the fact that in $\hat{\boldsymbol{p}}$ we estimate $D - 1$ parameters freely, but under H_0 we restrict p_i to be equal to one of the other $D - 1$ estimated parameters and hence only $D - 2$ parameters are estimated freely in $\boldsymbol{p}^*$. For $d \geq 3$ one has to recall that the maximisation is not resticted to the space where d p_j's are equal, but instead to the space where at least another p_j is equal to p_i, and hence an asymptotic χ^2 distribution with $d - 1$ degrees of freedom should not be expected.

To calculate the p-value of the observed λ one has to consider the probability $(1 - \delta)$ of obtaining a positive test statistic. A simple estimate of δ is $\hat{\delta} = (d-1)/d$, where d is the number of estimates in $\boldsymbol{p}^*$ equal or almost equal to p_i^* (including p_i^*). An approximate p-value of the test may then be obtained as

$$\frac{1 - F(d\lambda/2)}{d},$$

Table 1: The reported estimated voter shares $\hat{p}$, the corresponding frequencies x (assuming a simple random sample), and the estimated voter shares p^* under the restriction that SD are not allowed to be the largest party.

Party[a]	M	C	L	KD	MP	S	V	FI	SD	Other
$\hat{p}$	0.210	0.056	0.044	0.037	0.064	0.234	0.068	0.028	0.252	0.007
x	321	85	67	56	98	357	104	43	385	11
p^*	0.210	0.055	0.043	0.036	0.064	0.244	0.070	0.029	0.244	0.007

[a]The Moderates (M), the Centre Party (C), the Liberals (L), the Christian Democrats (KD), the Green Party (MP), the Social Democrats (S), the Left Party (V), the Feminist Party (FI), and the Sweden Democrats (SD).

where F is the cumulative density function of the χ^2-distribution with one degree of freedom.

Simulations indicate that the asymptotic properties work well, see Appendix B for details.

4. An example

As an example we use the poll that inspired this paper. The poll that *Metro* published was done by YouGov Sweden. In total 1527 people responded to the poll. Nine parties were reported yielding the estimates $\hat{p}$ in Table 1. We note that the three largest parties, the Moderates (M), the Social Democrats (S), and the Sweden Democrats (SD), are roughly equal in size: 21–25%. Since the reported shares of the nine parties sum to 0.993, the shares of all other parties must sum to 0.007.

YouGov used a self-recruited on-line panel for the poll, i.e. not a random sample from the electorate, and most likely weighted the answers in some intricate way. However, we will assume that the estimates are based on a simple random sample from the electorate. Given a sample of 1527 respondents, the reported shares would correspond to the frequencies x in Table 1.

The ML estimate p^* in the restricted parameter space, i.e. the space where SD are not the largest party, is given in Table 1. We note that the estimated shares of S and SD are equal, as p^* is restricted to $\mathbb{S}^{10} \setminus \omega_{SD}$ including its boundary. The estimate p^* corresponds to a log likelihood value of $\log L(p^*) = -28.038$, whereas the log likelihood value for the ML estimate in the unrestricted parameter space is $\log L(\hat{p}) = -27.452$. This gives a test statistic of $\lambda = 1.171$, yielding a p-value between $(1 - F(2 \cdot 1.171))/2 = 0.14$ and $(1 - F(3 \cdot 1.171))/3 = 0.06$. Based on the data, we would not draw the conclusion that the Sweden Democrats are the largest party in Sweden at a five per cent level of significance.

5. Concluding remarks

We have introduced a novel maximum likelihood ratio test for testing if a specific part or proportion is the largest among D parts. The test is very general and can be applied to any situation, with a finite number of parts, where one wants to test if the observed frequencies support the hypothesis that a specific share of the population is the largest. The test is fairly straightforward and attains a specified size. In Appendix B we present a small simulation study indicating that the empirical size of the test is reasonably close to the theoretical size; in our simulations the empirical sizes varied from 0.030 to 0.055 compared with the theoretical size 0.05. An important aspect of the test is that it is based on, and respects, the parameter space of the problem.

It remains as future research to develop an exact expression for δ as a function of p^*. However, at this point we believe that such expressions would be of more theoretical value than of dramatically improved practical usefulness.

We hope that this test can be of practical use for many, not only when analysing political poll results. It provides a simple and stringent way of testing a statement of great interest to many people.

Appendix A. Two theorems with proofs

Theorem 1. *If $p \in \mathbb{S}^D$ and p_i is the largest part, then $p_i > 1/D$.*

Table B.2: Proportion of test statistics equal to zero and the empirical size for different combinations of $\boldsymbol{p}$ and sample size.

	Proportion of 0's					Empirical size				
$\boldsymbol{p}$	300	600	900	1200	1500	300	600	900	1200	1500
1	0.521	0.507	0.519	0.511	0.503	0.049	0.049	0.030	0.039	0.051
2	0.674	0.687	0.683	0.677	0.680	0.044	0.049	0.040	0.052	0.043
3	0.819	0.804	0.799	0.827	0.828	0.051	0.049	0.054	0.042	0.048
4	0.568	0.518	0.532	0.522	0.503	0.030	0.037	0.038	0.043	0.055

PROOF. Let p_i be the largest part. If $p_i < 1/D$, then all other parts are also less than $1/D$, and the sum of all parts is then less than 1. If $p_i = 1/D$ then either $p_j < 1/D$ for some $j \neq i$ and the sum of all parts is less than 1, or $p_j = 1/D$ for all $j = 1, \ldots, D$ but then p_i is not the largest part. □

Theorem 2. *If $\boldsymbol{p} \in \mathbb{S}^D$ and $p_i > 1/2$, then p_i is the largest part.*

PROOF. If p_i is not the largest part and $p_i > 1/2$, then there exists a part $p_j > p_i$ for some $j \neq i$. But then also $p_j > 1/2$ and the sum of the parts is greater than 1. □

Appendix B. Simulation study of the distribution of the test statistic

As an example we present a small simulation to verify the properties of the test statistic. Four different $\boldsymbol{p}$ were used:

$$1: \quad \boldsymbol{p} = (0.48, 0.48, 0.04)'$$
$$2: \quad \boldsymbol{p} = (1/3, 1/3, 1/3)'$$
$$3: \quad \boldsymbol{p} = (0.15, 0.15, 0.04, 0.05, 0.06, 0.15, 0.07, 0.03, 0.15, 0.15)'$$
$$4: \quad \boldsymbol{p} = (0.24, 0.06, 0.04, 0.04, 0.06, 0.24, 0.07, 0.03, 0.21, 0.01)'$$

In the first $\boldsymbol{p}$ two parts are equal and the third much, much smaller; in the second all three parts are equal; in the third five out of ten parts are equal; the fourth $\boldsymbol{p}$ resembles the $\boldsymbol{p}^*$ in the example. For each $\boldsymbol{p}$ 1000 random samples were drawn from a multinomial distribution using sample sizes $n = 300, 600, 900, 1200,$ and 1500. For each sample we tested $H_0 : \boldsymbol{p} \in \omega_1$. Figure B.2 shows the QQ-plot of $d\lambda/2$ (for non-zero test statistics) for the different combinations of $\boldsymbol{p}$ and sample size n compared with the $\chi^2(1)$-distribution. Table B.2 presents the proportion of test statistics equal to zero and the empirical size of the test, i.e. the proportion of 1000 statistics with a p-value less than 0.05. We conclude that the test statistics seem to be reasonably χ^2-distributed even for small sample sizes. The proportion of zeros is also close the estimated number $(d-1)/d$. The empirical sizes are close to the nominal 0.05.

References

Aitchison, J., 1982. The statistical analysis of compositional data. J. R. Stat. Soc. Ser. B Stat. Methodol. 44 (2), 139–177.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C., 2003. Isometric logratio transformations for compositional data analysis. Math. Geol. 35 (3), 279–300.

Garthwaite, P. H., Jolliffe, I. T., Jones, B., 2002. Statistical Inference. Oxford University Press, Oxford.

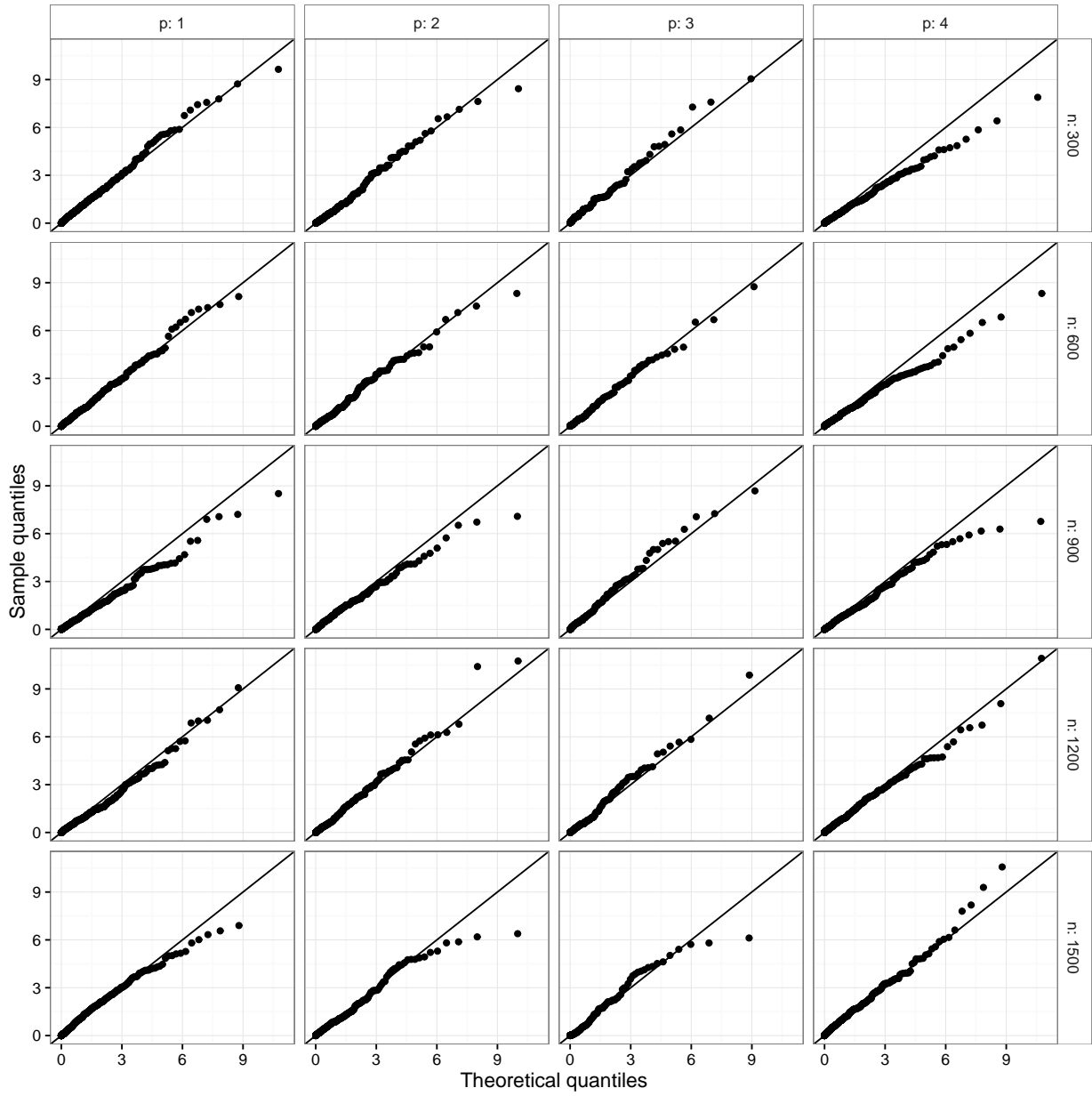Wallroth, E., 2015. Nu är SD Sveriges största parti. Metro Skåne Edition. 20 August 2015.

Figure B.2: QQ-plots of simulated non-zero test statistics for $n = 300, 600, 900, 1200, 1500$. The different $\boldsymbol{p}$ used are 1: $(0.48, 0.48, 0.04)'$, 2: $(1/3, 1/3, 1/3)'$, 3: $(0.15, 0.15, 0.04, 0.05, 0.06, 0.15, 0.07, 0.03, 0.15, 0.15)'$, and 4: $(0.24, 0.06, 0.04, 0.04, 0.06, 0.24, 0.07, 0.03, 0.21, 0.01)'$. (In the bottom left panel one extreme observation has been excluded for enhanced clarity.)