Universality of neutral models: Decision process in politics

Supplementary material. Mathematical analysis and data.

December 6, 2018

1 Overview over the mathematical analysis

1.1 Data driven modeling

We summarize the considerations made in the main text. The starting point of our considerations is the observation that the number of votes per party satisfies approximately a log-linear dependency on the party's rank. The slope of this log-linear relation seems to be rather independent of the number of voters: If we inspect the parameters of the linear model, we only find the intercept to vary with the size of the group. The slope is fairly constant. That is, we expect that the mechanism is the same on different levels of organization (city, state, country).
The decision process leads to group formation – we identify the voters of a given party with one group. This process inherits, of course, stochasticity. That is, if we consider any partition of the population, the resulting decomposition has a certain probability to meet the "true" decomposition, observed in elections. As the data indicate that the slope is basically independent on the size of the organizational unit, we expect that the fundamental ruling mechanisms are fairly independent of the population size (number of voters). Even more, if we first take a subset of the complete population (the voters in Bavaria, say), and the a subset of the subset (the voters in Munich), yields the same result as if we directly take the small sample (Munich). In statistics, this property is known as sampling consistency condition [9, 8]. The probability measures that exhibit the sampling consistency condition are well characterized: they can be constructed by the so-called paint ball process [9]. One distribution out of those is especially famous: the Ewens sampling distribution [5, 4]. This distribution has many applications, in particular in population genetics [3], and moreover, can be generated by a stochastic process – the infinite allele Moran model with mutations.

We adapt the infinite allele model in the context of opinion formation of voters (see Figure 1). Let us consider a population of n voters – non-voters are neglected. Each voter is a supporter of a proto-party. The difference between a party and a proto-party becomes clear below, by now we may identify the two terms. Voters change their opinion in the following way: a randomly selected voter thinks over his/her opinion. With a certain probability v, he/she stays with his/her opinion. With probability $1 - v$ the person is prepared to change the proto-party he/she is supporting. If this is the case, this person either constitutes a new proto-party with probability u, or selects randomly one person of the population and adopts the opinion of that individual. If the last supporting voter of a given proto-party changes his/her mind such that this proto-party has no supporters any more, this proto-party is dissolved. In particular, no political aims or believes are involved in this model – it is a model neutral.

In the long run, the stochastic process approaches its invariant measure. There is some kind of equilibrium for the number of proto-parties (a distribution), and each individual is the follower of a proto-party. In an election, however, proto-parties that are too small will not stand for election. Only

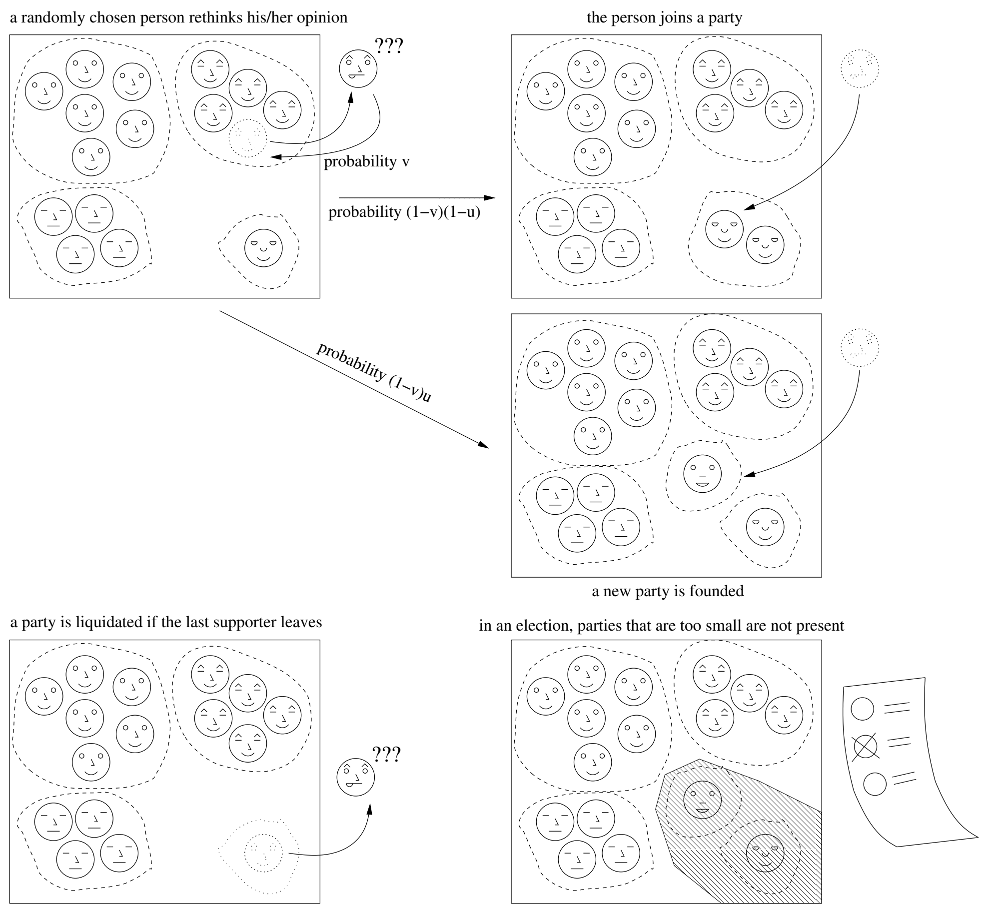


Figure 1: Scheme of the model.

proto-parties with a supercritical number of followers n_0 become parties. It turns out later that the relative critical size $z = n_0/n$ is even more handy parameter. Followers of a subcritical proto-party give their vote to another party, where that party is selected proportional to its size. We call this model the voter model with party-dynamics.

The voter model with party dynamics and $n_0 = 0$ is equivalent with an infinite-allele Moran model with mutation. It is well known that the resulting composition structure follows the Ewens sampling formula, which has only one parameter $\theta = 2\,u\,n/(1-v)$, the rescaled party-constitution rate (or the rescaled mutation rate in the context of population genetics). Our model has one additional parameter: the threshold $z \in (0, 1)$. All individuals in groups with a size smaller $n_0 = z\,n$ are distributed to the groups with size larger or equal $z\,n$.

2 Invariant measure of the model, Ewens Sampling Formula

2.1 The Ewens Sampling Formula

We recall some basic facts about the Ewens Sampling foprmula. The Ewens Sampling Formula is a probability measure that describes partitions of a set (population) of size n. The Ewens Sampling

Formula is appealing in that this probability measure is generated by the invariant measure of the infinite allele model, that is, of a mechanistic and interpretable process. If $z_0 = 0$, this process is identical with our voter model with party dynamics.
Let us first consider the original Ewens Sampling Formula. In a sample of size n (given), we observe K groups (K is a random variable). The sizes are given by $a_1, \ldots, a_K$. Let c_1 denote the number of groups of size 1, c_2 the number of groups of size 2, etc. The partition can be characterized by the vector of frequencies $C = (c_1, \ldots c_n)$, where $n = \sum_{i=1}^n i\, c_i$ and $K = \sum_{i=1}^n c_i$. The Ewens Sampling Formula states the probability to observe a given partition structure C. Mostly, it is parametrized by a scaled mutation probability θ, and can be formulated as (e.g. [4, page 22])

$$C \sim (Y_1, \ldots Y_n | \sum iY_i = n) \tag{1}$$

where $Y_i \sim \text{Pois}(\theta/i)$ are independent Poisson random variables.

In our application it will be interesting to condition the distribution on the number of groups [4, chapter 3.1], [11], [6, chapter 9.5]. Note that the number of groups $K = K(C) = \sum_i c_i$ is a random variable. Then,

$$P(K = k|\theta) = \frac{\theta^k}{\theta_{(n)}}\, |S_n^k| \tag{2}$$

where S_n^k are the number of permutations of $\{1, \ldots, n\}$ with exactly k cycles. The Ewens Sampling Formula reads ($k = \sum c_j$ in the given realization)

$$P(C|\theta) = \frac{n!}{\theta_{(n)}} \prod_{j=1}^{n} \frac{(\theta/j)^{c_j}}{c_j!} = \frac{n!\,\theta^k}{\theta_{(n)}} \prod_{j=1}^{n} \frac{(1/j)^{c_j}}{c_j!}.$$

Therefore,

$$P(C|\theta, K = k) \quad = \quad \frac{P(c_1, \ldots, c_n \quad \text{and} \quad K = k|\theta)}{P(K = k|\theta)} = \frac{P(c_1, \ldots, c_n|\theta)}{P(K = k|\theta)} = \frac{n!}{|S_n^k|} \prod_{j=1}^{n} \frac{(1/j)^{c_j}}{c_j!}.$$

In a slight abuse of notation, we define for $a_i \geq 0$, $\sum a_i > 0$

$$\text{Multinom}(n, (a_1, \ldots, a_m))$$

to denote the multinomial distribution with n trials, and $p_i = a_i / \sum_{j=1}^m a_j$. If we compare the result above with the probability function of a multinomial distribution, we conclude

$$(C|K = k) \;\sim \text{Multinom}(k, (1, 1/2, \ldots, 1/n)) \;\Big|\; \sum_{i=1}^{n} ic_i = n. \tag{3}$$

2.2 Invariant measure of the model

In our model we know the number of parties before the election takes place. That is, the natural parametrization of our model uses the minimal group size possible $n_0 = \lfloor z\, n \rfloor$, the population size n, and the number of observed groups K_+ (that is, groups with a supercritical size, $a_i \geq \lfloor z\, n \rfloor$). The aim of the present section is to show that, conditioned on these parameters, the invariant measure of our model can be well approximately by a multinomial distribution.

In our case, we are only interested in groups with size above the critical group size n_0. For the random variable $C = (c_1, \dots, c_n)$ define

$$K_+ = K_+(C) = \sum_{i=n_0}^{n} c_i, \quad N_+ = N_+(C) = \sum_{i=n_0}^{n} i\, c_i, \quad K_- = \sum_{i=1}^{n_0-1} c_i = K - K_+, \quad N_+ = \sum_{i=1}^{n_0-1} i\, c_i = n - N_+.$$

Furthermore, we define the projections

$$\Pi_+(c_1, \dots, c_n) = (0, \dots, 0, c_{n_0}, \dots, c_n), \quad \Pi_-(c_1, \dots, c_n) = (c_1, \dots, c_{n_0-1}, 0 \dots, 0).$$

Our ultimate goal is an approximation of the distribution of $\Pi_+(C)|\theta, K_+$. If we do not only condition on K and K_+, but also on N_+, then

$$(\Pi_+(C)|K = k,\ K_+ = k_+,\ N_+ = n_+) \sim \text{Multinom}\Big(k_+, (0, \dots, 0, 1/n_0, \dots, 1/n)\Big) \ \Big|\ \sum_{i=n_0}^{n} i c_i = n_+.$$

Note that the r.h.s. does not depend on K any more. Thus, we may drop the condition on K,

$$(\Pi_+(C)|K_+ = k_+,\ N_+ = n_+) \sim \text{Multinom}\Big(k_+, (0, \dots, 0, 1/n_0, \dots, 1/n)\Big) \ \Big|\ \sum_{i=n_0}^{n} i c_i = n_+.$$

If $K_+ = k_+$ is fixed, we have a hierarchical model: We obtain a conditioned multinomial distribution for $\Pi_+(C)|\theta,\ K_+ = k_+,\ N_+ = n_+$, where θ does not appear directly any more; however, the distribution of N_+ depends on θ. That is, n_+ can be considered as a hyperparameter with a given (θ-dependent) distribution. In that, the situation is somewhat similar to that before: if we observe all groups, the distribution of $C|\theta, K$ only depends on θ via the distribution of K. Also (3) can be interpreted as an hierarchical model with hyperparameter K, which has the distribution indicated in (2).

In order to infer the distribution of N_+, we note $N_+ = n - N_-$. Next we use (see [4, Theorem 1.19], [1]) that for n large, approximately $c_i \sim Y_i$, where $Y_i \sim \text{Pois}(\theta/i)$ are independent random variables. Note that in this step some heuristics are involved, as this theorem is only true for i fixed and $n \to \infty$. For n large we have approximately

$$N_- \sim \sum_{i=1}^{n_0-1} Y_i \sim \text{Pois}(\sum_{i=1}^{n_0-1} \theta/i).$$

Note that this Poisson random variable may attend arbitrary large numbers, such that in this approximation $N_+ \leq n$ is not always given. However,

$$E(N_-) \approx \sum_{i=1}^{n_0-1} \theta/i \approx \theta \log(n_0).$$

This argument indicates that the hyperparameter N_+ can be represented as $n - N_-$, where N_- has a Poisson distribution with expectation $\mathcal{O}(\log(n_0))$. Thus, for n large, $N_+ \approx n$, and we find in this sense

$$(\Pi_+(C)|\theta,\ K_+ = k_+,\ N_+ = n) \approx \text{Multinom}(k_+, (0, \dots, 0, 1/n_0, \dots, 1/n)) \ \Big|\ \sum_{i=n_0}^{n} i c_i = n_+. \qquad (4)$$

This statistics can be reformulated in a handy way if the underlying urn model for a multinomial distribution is considered. We return to the group sizes $(a_1, \dots, a_{k_+})$; recall that the sample configuration

$C = (c_{n_0}, \dots, c_n)$ is computed from a realization of $(a_1, \dots, a_{k_+})$. We define i.i.d. random variables X_ℓ, $\ell \in \{1, \dots, k_+\}$, with values in $\{n_0, \dots, n\}$ where

$$P(X_\ell = i) = c/i, \tag{5}$$

and, as before,

$$c^{-1} = \sum_{i=n_0}^{n} i^{-1} \approx \ln((n+1)/n_0) \approx \ln(1/z), \tag{6}$$

then

$$(a_1, \dots a_{k_+})|\ N_+ = n \ \sim\ (X_1, \dots, X_{k_+}) \,\Bigg|\, \sum_{\ell=1}^{k_+} X_\ell = n, \tag{7}$$

The approximate distribution of configurations $(\Pi_+(C)|\theta,\ K_+ = k_+,\ N_+ = n)$ given in (4) is identical with the distributions of configurations generated by $(a_1, \dots a_{k_+})$ defined in (7). The invariant measure of our dynamic model, the voter model with party dynamics, is well described by the construction (7).

2.3 Rank statistics – unconditioned case

In this section, we derive an expression for the expectation of the size ratio of subsequent groups (ordered according to their rank), respectively about the expectation of the logarithm of the group size. Here we do not condition on the total population size. That is, we consider K independent realizations $X_1, \dots, X_K$ of i.i.d. random variables that assume values in $\{n_0, \dots, n\}$, where $0 < n_0 < n$, $P(X_i = j) = c/j$ for $j \in \{n_0, \dots, n\}$ and 0 else, $c^{-1} = \sum_{j=n_0}^{n} j^{-1}$. We order these realizations according to size $X_{(1)} \le X_{(2)} \dots \le X_{(K)}$, and investigate $E(X_{(\ell+1)}/X_{(\ell)})$ respectively $E(\ln(X_{(\ell+1)}))$. The central result of this section is the independence of that expectation w.r.t. ℓ for n large:

Theorem 2.1 *Let $\theta = K/\ln(1/z)$. For $\ell \in \{1, \dots, K-1\}$ we find*

$$\lim_{n\to\infty} E(X_{(\ell+1)}/X_{(\ell)}) \;=\; G(\theta, z) := 1 + \theta \int_z^1 \left(\frac{1}{z} - \frac{1}{y}\right) \left(\frac{\ln(1/y)}{\ln(1/z)}\right)^{\ln(1/z)\,\theta - 1} dy. \tag{8}$$

That is, the expectation of the quotient of subsequent group sizes is constant in the rank. The next proposition indicates that $E(X_{(\ell+1)}/X_{(\ell)})$ only depends weakly on z (for z small) if θ is larger one and kept fixed.

Proposition 2.2 *We find for $\theta > 1$ that*

$$\lim_{z\to 0} G(\theta, z) \;=\; \frac{\theta}{\theta - 1}. \tag{9}$$

For the logarithm, we obtain the following theorem.

Theorem 2.3 *For $\ell = 1, \cdots, K$, we find*

$$\begin{aligned} \lim_{n\to\infty} \quad & E(\ln(X_{(\ell+1)}) - \log(n)) \\ = \; & -\sum_{j=0}^{\ell-1} \binom{K}{j} \int_z^1 \frac{1}{x} \left(1 - \frac{\ln(1/x)}{\ln(1/z)}\right)^{j-1} \left(\frac{\ln(1/x)}{\ln(1/z)}\right)^{K-j} \left(K - j - K\frac{\ln(1/x)}{\ln(1/z)}\right) dx. \end{aligned} \tag{10}$$

This theorem indicates that the logarithm of the population size does not exactly depends in a linear way on the rank ℓ, but only approximatively.

We prove the theorems resp. proposition in the next sections.

2.3.1 Order statistics

As a first step we obtain $P(X_{(\ell)} = i)$, based on the well known formulas for the distribution functions of order statistics.

In the following, we use the convention that a sum extending from a to b with $a > b$ is zero, in particular

$$\sum_{i=n+1}^{n} (\cdots) := 0.$$

Proposition 2.4

$$\begin{aligned} P(X_{(1)} = i_1) &= \left(\sum_{j=i_1}^{n} \frac{c}{j}\right)^K - \left(\sum_{j=i_1+1}^{n} \frac{c}{j}\right)^K \\ P(X_{(\ell)} = i_1) &= \frac{c^K}{i_1} \sum_{j=0}^{\ell-1} \binom{K}{j} \Bigg\{ \left(\sum_{w=n_0}^{i_1-1} \frac{1}{w}\right)^{j-1} \left(\sum_{m=i_1+1}^{n} \frac{1}{m}\right)^{K-j-1} \left[K \left(\sum_{w=n_0}^{i_1-1} \frac{1}{w}\right) - jc^{-1}\right]\Bigg\} + \mathcal{O}(i_1^{-2}). \end{aligned}$$

Proof: Since $P(X_\ell \geq i_1) = \sum_{i=i_1}^{n} c/i$ for $\ell = 1, \ldots, K$, we have $P(X_{(1)} \geq i_1) = \left(\sum_{j=i_1}^{n} \frac{c}{j}\right)^K$ and

$$P(X_{(1)} = i_1) = P(X_{(1)} \geq i_1) - P(X_{(1)} \geq i_1 + 1) = \left(\sum_{j=i_1}^{n} \frac{c}{j}\right)^K - \left(\sum_{j=i_1+1}^{n} \frac{c}{j}\right)^K$$

Furthermore, we find for the ℓ'th order statistics $X_{(\ell)}$

$$\begin{aligned} P(X_{(\ell)} \geq i) &= P(\text{at most } \ell - 1 \text{ realizations are smaller } i) \\ &= \sum_{j=0}^{\ell-1} \binom{K}{j} P(X_1 < i)^j \; P(X_1 \geq i)^{K-j} = c^K \sum_{j=0}^{\ell-1} \binom{K}{j} \left(\sum_{w=n_0}^{i-1} \frac{1}{w}\right)^j \left(\sum_{m=i}^{n} \frac{1}{m}\right)^{K-j}. \end{aligned}$$

Hence, by means of Taylor expansion we obtain

$$\begin{aligned}
P(X_{(\ell)} = i_1) &= P(X_{(\ell)} \geq i_1) - P(X_{(\ell)} \geq i_1 + 1) \\
&= c^K \sum_{j=0}^{\ell-1} \binom{K}{j} \left\{ \left(\sum_{w=n_0}^{i_1-1} \frac{1}{w} \right)^j \left(\sum_{m=i_1}^{n} \frac{1}{m} \right)^{K-j} - \left(\sum_{w=n_0}^{i_1} \frac{1}{w} \right)^j \left(\sum_{m=i_1+1}^{n} \frac{1}{m} \right)^{K-j} \right\} \\
&= c^K \sum_{j=0}^{\ell-1} \binom{K}{j} \Bigg\{ \left(\sum_{w=n_0}^{i_1-1} \frac{1}{w} \right)^j \left(\sum_{m=i_1+1}^{n} \frac{1}{m} + \frac{1}{i_1} \right)^{K-j} - \left(\sum_{w=n_0}^{i_1-1} \frac{1}{w} + \frac{1}{i_1} \right)^j \left(\sum_{m=i_1+1}^{n} \frac{1}{m} \right)^{K-j} \Bigg\} \\
&= c^K \sum_{j=0}^{\ell-1} \binom{K}{j} \Bigg\{ \left(\sum_{w=n_0}^{i_1-1} \frac{1}{w} \right)^j \left[\left(\sum_{m=i_1+1}^{n} \frac{1}{m} \right)^{K-j} + \frac{K-j}{i_1} \left(\sum_{m=i_1+1}^{n} \frac{1}{m} \right)^{K-j-1} \right] \\
&\qquad - \left[\left(\sum_{w=n_0}^{i_1-1} \frac{1}{w} \right)^j + \frac{j}{i_1} \left(\sum_{w=n_0}^{i_1-1} \frac{1}{w} \right)^{j-1} \right] \left(\sum_{m=i_1+1}^{n} \frac{1}{m} \right)^{K-j} \Bigg\} + \mathcal{O}(i_1^{-2}) \\
&= \frac{c^K}{i_1} \sum_{j=0}^{\ell-1} \binom{K}{j} \Bigg\{ \left(\sum_{w=n_0}^{i_1-1} \frac{1}{w} \right)^{j-1} \left(\sum_{m=i_1+1}^{n} \frac{1}{m} \right)^{K-j-1} \\
&\qquad \times \left[(K-j) \left(\sum_{w=n_0}^{i_1-1} \frac{1}{w} \right) - j \left(\sum_{m=i_1+1}^{n} \frac{1}{m} \right) \right] \Bigg\} + \mathcal{O}(i_1^{-2}) \\
&= \frac{c^K}{i_1} \sum_{j=0}^{\ell-1} \binom{K}{j} \Bigg\{ \left(\sum_{w=n_0}^{i_1-1} \frac{1}{w} \right)^{j-1} \left(\sum_{m=i_1+1}^{n} \frac{1}{m} \right)^{K-j-1} \left[K \left(\sum_{w=n_0}^{i_1-1} \frac{1}{w} \right) - jc^{-1} \right] \Bigg\} + \mathcal{O}(i_1^{-2})
\end{aligned}$$

□

Therewith, we find

$$\begin{aligned}
E(X_{(1)}) &= \sum_{i=n_0}^{n} i \left(\left(\sum_{j=i}^{n} \frac{c}{j} \right)^K - \left(\sum_{j=i+1}^{n} \frac{c}{j} \right)^K \right) \\
&= \left(\sum_{i=n_0}^{n} i \left(\sum_{j=i}^{n} \frac{c}{j} \right)^K \right) - \left(\sum_{i=n_0}^{n} (i+1) \left(\sum_{j=i+1}^{n} \frac{c}{j} \right)^K \right) + \left(\sum_{i=n_0}^{n} \left(\sum_{j=i+1}^{n} \frac{c}{j} \right)^K \right) \\
&= n_0 + \left(\sum_{i=n_0}^{n} \left(\sum_{j=i+1}^{n} \frac{c}{j} \right)^K \right) \qquad (11)
\end{aligned}$$

where we used that $\sum_{j=n_0}^{n} c/j = 1$ and $\sum_{j=n+1}^{n}(\dots) = 0$.

Let us introduce some more notation. Denote by $X_{(\ell);n_0,n,K}$ the random variable as introduced above; the additional indices characterize all parameters of the random variable.

If we condition on $X_{(1)} = i_1$, then $X_{(2)}, \dots, X_{(K)}$ is the order statistics of $K-1$ random variables with values in $i_1, \dots, n$. That is, we obtain realizations $X_{(2)}, \dots, X_{(K)}$ by determining $K-1$ realizations of random variables Y_i with values in $i_1, \dots, n$, where $P(Y_i = j) = \tilde{c}_{i_1,n}/j$. Here, as before, $\tilde{c}_{i_1,n}^{-1} = \sum_{j=i_1}^{n} j^{-1}$. Then,

$$(X_{(2)}, \dots, X_{(K)}) | X_{(1)} = i_1 \;\sim\; (Y_{(1)}, \dots, Y_{(K-1)}).$$

In particular,

$$X_{(2);n_0,n,K}|X_{(1);n_0,n,K}=i_1 \quad \sim \quad X_{(1);i_1,n,K-1}$$

and, similarly, for $\ell = 1,\dots,K-1$,

$$X_{(\ell+1);n_0,n,K}|X_{(\ell);n_0,n,K}=i_1 \quad \sim \quad X_{(1);i_1,n,K-\ell} \tag{12}$$

2.3.2 Size ratio

Before we come to the point where we investigate $\lim_{n\to\infty} E(X_{(\ell+1)}/X_{(\ell)})$, we first indicate two algebraic relations. All computations below are not deep, but lengthy and involving. We do all calculations step by step, even the completely trivial parts that are most likely unnecessary for trained readers.

Proposition 2.5

$$\sum_{m=0}^{\ell-1}(-1)^m\binom{K}{\ell-1-m}\binom{\ell-(\ell-1-m)-1}{m}K=\frac{1}{(\ell-1)!}\prod_{j=0}^{\ell-1}(K-j).$$

Proof: This proposition is a consequence of Pascal's identity. First we note that

$$\frac{1}{(\ell-1)!}\prod_{j=0}^{\ell-1}(K-j)=\frac{K!}{(\ell-1)!(K-\ell)!}=\frac{K!}{(\ell-1)!(K-1-(\ell-1))!}=K\binom{K-1}{\ell-1}$$

Then,

$$\begin{aligned}
&\sum_{m=0}^{\ell-1}(-1)^m\binom{K}{\ell-1-m}\binom{\ell-(\ell-1-m)-1}{m}K\\
=\;&K\sum_{m=0}^{\ell-1}(-1)^m\binom{K}{\ell-1-m}=K\sum_{m=0}^{\ell-2}(-1)^m\left[\binom{K-1}{\ell-1-m-1}+\binom{K-1}{\ell-1-m}\right]+(-1)^{\ell-1}\binom{K}{0}K\\
=\;&K\left[\binom{K-1}{\ell-2}+\binom{K-1}{\ell-1}\right]\\
&+K\sum_{m=2}^{\ell-1}(-1)^{m+1}\binom{K-1}{\ell-1-m}+K\sum_{m=1}^{\ell-2}(-1)^m\binom{K-1}{\ell-1-m}+(-1)^{\ell-1}K\\
=\;&K\left[\binom{K-1}{\ell-2}+\binom{K-1}{\ell-1}\right]+K(-1)^{\ell}\binom{K-1}{0}-K\binom{K-1}{\ell-2}+(-1)^{\ell-1}K=K\binom{K-1}{\ell-1}
\end{aligned}$$

□

Proposition 2.6 *For $n\in\{0,\dots,\ell-2\}$,*

$$\sum_{m=0}^{n}(-1)^m\left\{\binom{K}{n-m}\binom{\ell-(n-m)-1}{m}K-\binom{K}{n+1-m}(n+1-m)\binom{\ell-(n+1-m)-1}{m}\right\}=0$$

Proof: With

$$\binom{K}{n+1-m}(n+1-m)=\binom{K-1}{n-m}K$$

we find

$$
\begin{aligned}
& \sum_{m=0}^{n}(-1)^m \binom{K}{n-m} K \binom{\ell-(n-m)-1}{m} \\
& -\sum_{m=0}^{n}(-1)^m \binom{K}{n+1-m} (n+1-m) \binom{\ell-(n+1-m)-1}{m} \\
= \;& K \sum_{m=0}^{n}(-1)^m \left\{ \binom{K}{n-m} \binom{\ell-(n-m)-1}{m} - \binom{K-1}{n-m} \binom{\ell-(n-m)-2}{m} \right\} \\
= \;& K \sum_{m=0}^{n}(-1)^{n-m} \left\{ \binom{K}{m} \binom{\ell-m-1}{n-m} - \binom{K-1}{m} \binom{\ell-m-2}{n-m} \right\} \\
= \;& K(-1)^n \binom{\ell-1}{n} + K \sum_{m=1}^{n}(-1)^{n-m} \left\{ \binom{K}{m} \binom{\ell-m-1}{n-m} \right\} \\
& -K \sum_{m=0}^{n-1}(-1)^{n-m} \left\{ \binom{K-1}{m} \binom{\ell-m-2}{n-m} \right\} - K \binom{K-1}{n} \\
= \;& K(-1)^n \binom{\ell-1}{n} - K \sum_{m'=0}^{n-1}(-1)^{n-m'} \left\{ \binom{K}{m'+1} \binom{\ell-m'-2}{n-m'-1} \right\} \\
& -K \sum_{m=0}^{n-1}(-1)^{n-m} \left\{ \binom{K-1}{m} \binom{\ell-m-2}{n-m} \right\} - K \binom{K-1}{n} \\
= \;& K(-1)^n \binom{\ell-1}{n} - K \sum_{m=0}^{n-1}(-1)^{n-m} \left\{ (\binom{K-1}{m+1} + \binom{K-1}{m}) \binom{\ell-m-2}{n-m-1} \right\} \\
& -K \sum_{m=0}^{n-1}(-1)^{n-m} \left\{ \binom{K-1}{m} \binom{\ell-m-2}{n-m} \right\} - K \binom{K-1}{n} \\
= \;& K(-1)^n \binom{\ell-1}{n} - K \sum_{m=0}^{n-1}(-1)^{n-m} \left\{ \binom{K-1}{m+1} \binom{\ell-m-2}{n-m-1} \right\} \\
& -K \sum_{m=0}^{n-1}(-1)^{n-m} \left\{ \binom{K-1}{m} (\binom{\ell-m-2}{n-m} + \binom{\ell-m-2}{n-m-1} \right\} - K \binom{K-1}{n} \\
= \;& K(-1)^n \binom{\ell-1}{n} - K \sum_{m=0}^{n-1}(-1)^{n-m} \left\{ \binom{K-1}{m+1} \binom{\ell-(m+1)-1}{n-(m+1)} \right\} \\
& -K \sum_{m=0}^{n-1}(-1)^{n-m} \left\{ \binom{K-1}{m} \binom{\ell-m-1}{n-m} \right\} - K \binom{K-1}{n}
\end{aligned}
$$

$$
\begin{aligned}
= \;& K(-1)^n \binom{\ell-1}{n} + K \sum_{m=1}^{n}(-1)^{n-m} \left\{ \binom{K-1}{m} \binom{\ell-m-1}{n-m} \right\} \\
& -K \sum_{m=0}^{n-1}(-1)^{n-m} \left\{ \binom{K-1}{m} \binom{\ell-m-1}{n-m} \right\} - K \binom{K-1}{n} = 0.
\end{aligned}
$$

□

Proposition 2.7

$$\lim_{n\to\infty} E(X_{(\ell+1)}/X_{(\ell)}) - 1$$
$$= \frac{1}{(\ell-1)!}\prod_{j=0}^{\ell-1}(K-j)\int_0^{1/z-1}\left(1-\frac{\ln(u+1)}{\ln(1/z)}\right)^{K-\ell}\int_1^{1+u}\frac{\ln(v)^{\ell-1}\ln(1/z)^{-\ell}}{v^2}\,dv\,du \qquad (13)$$

Proof: As $X_{(\ell+1);n_0,n,K}|X_{(\ell);n_0,n,K}=i_1 \sim X_{(1);i_1,n,K-\ell}$, we have $(c_{i,n} := (\sum_{h=i}^{n} h^{-1})^{-1})$

$$\begin{aligned}
E(X_{(\ell+1)}/X_{(\ell)}) &= E(E(X_{(\ell+1)}/X_{(\ell)}|X_{(\ell)})) = \sum_{i_1=n_0}^{n}\frac{1}{i_1}E(X_{(\ell+1)}|X_{(\ell)}=i_1)P(X_{(\ell)}=i_1)\\
&= \sum_{i_1=n_0}^{n}\frac{1}{i_1}E(X_{(1);i_1,n,K-\ell})\ P(X_{(\ell)}=i_1)\\
&= \sum_{i_1=n_0}^{n}\frac{1}{i_1}\left[i_1 + \left(\sum_{i=i_1}^{n}\left(\sum_{j=i+1}^{n}\frac{c_{i_1,n}}{j}\right)^{K-\ell}\right)\right]P(X_{(\ell)}=i_1)\\
&= 1+\sum_{i_1=n_0}^{n}\frac{1}{i_1}\left(\sum_{i=i_1}^{n}\left(\sum_{j=i+1}^{n}\frac{c_{i_1,n}}{j}\right)^{K-\ell}\right)P(X_{(\ell)}=i_1).
\end{aligned}$$

Hence, with proposition 2.4,

$$\begin{aligned}
&E(X_{(\ell+1)}/X_{(\ell)}) - 1\\
&= \sum_{j=0}^{\ell-1}\sum_{i_1=n_0}^{n}\sum_{i=i_1}^{n}\frac{c^K}{(i_1/n)^2}\left(\sum_{j'=i+1}^{n}\frac{c_{i_1,n}}{j'}\right)^{K-\ell}\left[\binom{K}{j}\left\{\left(\sum_{w=n_0}^{i_1-1}\frac{1}{w}\right)^{j-1}\left(\sum_{m=i_1+1}^{n}\frac{1}{m}\right)^{K-j-1}\times\right.\right.\\
&\qquad\left.\left.\times\left(K\sum_{w=n_0}^{i_1-1}\frac{1}{w}-jc^{-1}\right)\right\}+\frac{1}{n}\mathcal{O}((i_1/n)^{-3})\right]\frac{1}{n^2}.
\end{aligned}$$

Introducing $y=i/n\in[w^{-1},1]$, $w=i_1/n\in[z,1]$, we find for n large

$$\begin{aligned}
\sum_{j'=i+1}^{n}\frac{c_{i_1,n}}{j'} &= \frac{\ln((n+1)/i)}{\ln((n+1)/i_1)}+\mathcal{O}(i_1^{-1}) = \frac{\ln(1/y)}{\ln(1/w)}+\mathcal{O}(n^{-1})\\
\sum_{w=n_0}^{i_1-1}\frac{1}{w} &= \ln((i_1-1)/n_0)+\mathcal{O}(i_1^{-1}) = \ln\left(\frac{(i_1-1)/n)}{(n_0/n)}\right)+\mathcal{O}(i_1^{-1}) = \ln(w/z)+\mathcal{O}(n^{-1})\\
\sum_{m=i_1+1}^{n}\frac{1}{m} &= \ln(n/(i_1+1))+\mathcal{O}(i_1^{-1}) = \ln(1/w)+\mathcal{O}(n^{-1})
\end{aligned}$$

As $n_0=\lfloor z\,n\rfloor$, we note furthermore (see eqn. (6)) that $\lim_{n\to\infty}c^{-1}=\lim_{n\to\infty}\ln(n/n_0)=\ln(1/z)$. In the equation $E(X_{(\ell+1)}/X_{(\ell)})-1$ we recognize two nested Riemann sums; for $n\to\infty$, we find that these terms converge to nested integrals

$$
\begin{aligned}
&\lim_{n\to\infty} E(X_{(\ell+1)}/X_{(\ell)}) - 1 \\
= \; & \sum_{j=0}^{\ell-1} \int_z^1 \int_w^1 \frac{\ln(1/z)^{-K}}{w^2} \left(\frac{\ln(1/y)}{\ln(1/w)}\right)^{K-\ell} \binom{K}{j} \ln(w/z)^{j-1} \; \ln(1/w)^{K-j-1} \Big(K \, \ln(w/z) - j \ln(1/z) \Big) \, dy \, dw \\
= \; & \sum_{j=0}^{\ell-1} \binom{K}{j} K \, \ln(1/z)^{-K} \int_z^1 \ln(1/y)^{K-\ell} \int_z^y \frac{1}{w^2} \, \frac{\ln(w/z)^{j} \; \ln(1/w)^{K-j-1}}{\ln(1/w)^{K-\ell}} \, dw \; dy \\
& - \sum_{j=0}^{\ell-1} \binom{K}{j} j \, \ln(1/z)^{-K+1} \int_z^1 \ln(1/y)^{K-\ell} \int_z^y \frac{1}{w^2} \, \frac{\ln(w/z)^{j-1} \; \ln(1/w)^{K-j-1}}{\ln(1/w)^{K-\ell}} \, dw \; dy \\
= \; & \sum_{j=0}^{\ell-1} \binom{K}{j} K \, \ln(1/z)^{-K} \int_z^1 \ln(1/y)^{K-\ell} \int_z^y \frac{1}{w^2} \; \ln(w/z)^{j} \; \ln(1/w)^{\ell-j-1} \, dw \; dy \\
& - \sum_{j=0}^{\ell-1} \binom{K}{j} j \, \ln(1/z)^{-K+1} \int_z^1 \ln(1/y)^{K-\ell} \int_z^y \frac{1}{w^2} \; \ln(w/z)^{j-1} \; \ln(1/w)^{\ell-j-1} \, dw \; dy
\end{aligned}
$$

Now we transform the integrals: Let $y = z(u+1)$, then

$$
\begin{aligned}
& E(X_{(\ell+1)}/X_{(\ell)}) \\
= \; & 1 + \sum_{j=0}^{\ell-1} \binom{K}{j} K \, \ln(1/z)^{-K} \int_0^{1/z-1} \ln\left(\frac{1}{z(u+1)}\right)^{K-\ell} z \int_z^{z(1+u)} \frac{1}{w^2} \; \ln(w/z)^{j} \; \ln(1/w)^{\ell-j-1} \, dw \; du \\
& - \sum_{j=0}^{\ell-1} \binom{K}{j} j \, \ln(1/z)^{-K+1} \int_0^{1/z-1} \ln\left(\frac{1}{z(u+1)}\right)^{K-\ell} z \int_z^{z(1+u)} \frac{1}{w^2} \; \ln(w/z)^{j-1} \; \ln(1/w)^{\ell-j-1} \, dw \; du.
\end{aligned}
$$

And next, let $v = w/z$,

$$
\begin{aligned}
& E(X_{(\ell+1)}/X_{(\ell)}) \\
= \; & 1 + \sum_{j=0}^{\ell-1} \binom{K}{j} K \, \ln(1/z)^{-K} \int_0^{1/z-1} \ln\left(\frac{1}{z(u+1)}\right)^{K-\ell} \int_1^{1+u} \frac{1}{v^2} \; \ln(v)^{j} \; \ln(1/(z\,v))^{\ell-j-1} \, dv \; du \\
& - \sum_{j=0}^{\ell-1} \binom{K}{j} j \, \ln(1/z)^{-K+1} \int_0^{1/z-1} \ln\left(\frac{1}{z(u+1)}\right)^{K-\ell} \int_1^{1+u} \frac{1}{v^2} \; \ln(v)^{j-1} \; \ln(1/(z\,v))^{\ell-j-1} \, dv \; du \\
= \; & 1 + \sum_{j=0}^{\ell-1} \binom{K}{j} K \, \ln(1/z)^{-\ell} \int_0^{1/z-1} \left(1 - \frac{\ln(u+1)}{\ln(1/z)}\right)^{K-\ell} \int_1^{1+u} \frac{1}{v^2} \; \ln(v)^{j} \; \ln(1/(z\,v))^{\ell-j-1} \, dv \; du \\
& - \sum_{j=1}^{\ell-1} \binom{K}{j} j \, \ln(1/z)^{-\ell+1} \int_0^{1/z-1} \left(1 - \frac{\ln(u+1)}{\ln(1/z)}\right)^{K-\ell} \int_1^{1+u} \frac{1}{v^2} \; \ln(v)^{j-1} \; \ln(1/(z\,v))^{\ell-j-1} \, dv \; du.
\end{aligned}
$$

We expand $\ln(1/(zv))^{\ell-j-1} = [-\ln(v) + \ln(1/z)]^{\ell-j-1}$, and collect terms with equal powers of $\ln(v)$ and $\ln(1/z)$: If we use the abbreviation

$$
A = \left(1 - \frac{\ln(u+1)}{\ln(1/z)}\right)
$$

we have

$$
\begin{aligned}
&E(X_{(\ell+1)}/X_{(\ell)}) - 1 \\
= \; & \sum_{j=0}^{\ell-1} \binom{K}{j} K \, \ln(1/z)^{-\ell} \int_0^{1/z-1} A^{K-\ell} \int_1^{1+u} \frac{1}{v^2} \; \ln(v)^j \; \ln(1/(z\,v))^{\ell-j-1} \, dv \; du \\
& - \sum_{j=1}^{\ell-1} \binom{K}{j} j \, \ln(1/z)^{-\ell+1} \int_0^{1/z-1} A^{K-\ell} \int_1^{1+u} \frac{1}{v^2} \; \ln(v)^{j-1} \; \ln(1/(z\,v))^{\ell-j-1} \, dv \; du \\
= \; & \sum_{j=0}^{\ell-1} \binom{K}{j} K \, \ln(1/z)^{-\ell} \times \\
& \times \int_0^{1/z-1} A^{K-\ell} \int_1^{1+u} \frac{1}{v^2} \left(\sum_{m=0}^{\ell-j-1} \binom{\ell-j-1}{m} (-1)^m \ln(v)^m \; \ln(1/z)^{\ell-j-1-m} \right) \ln(v)^j \, dv \; du \\
& - \sum_{j=1}^{\ell-1} \binom{K}{j} j \, \ln(1/z)^{-\ell+1} \times \\
& \times \int_0^{1/z-1} A^{K-\ell} \int_1^{1+u} \frac{1}{v^2} \left(\sum_{m=0}^{\ell-j-1} \binom{\ell-j-1}{m} (-1)^m \ln(v)^m \; \ln(1/z)^{\ell-j-1-m} \right) \ln(v)^{j-1} \, dv \; du \\
= \; & \sum_{j=0}^{\ell-1} \binom{K}{j} K \int_0^{1/z-1} A^{K-\ell} \int_1^{1+u} \frac{1}{v^2} \left(\sum_{m=0}^{\ell-j-1} \binom{\ell-j-1}{m} (-1)^m \ln(v)^{m+j} \; \ln(1/z)^{-j-1-m} \right) dv \; du \\
& - \sum_{j=0}^{\ell-1} \binom{K}{j} j \int_0^{1/z-1} A^{K-\ell} \int_1^{1+u} \frac{1}{v^2} \left(\sum_{m=0}^{\ell-j-1} \binom{\ell-j-1}{m} (-1)^m \ln(v)^{m+j-1} \; \ln(1/z)^{-j-m} \right) dv \; du
\end{aligned}
$$

By $n = m + j$, we re-order the sums, noting that

$$
\sum_{j=0}^{\ell-1} \sum_{m=0}^{\ell-j-1} \mathrm{term}(j,m) = \sum_{n=0}^{\ell-1} \sum_{m=0}^{n} \mathrm{term}(n-m,m)
$$

Therewith,

$$
\begin{aligned}
&E(X_{(\ell+1)}/X_{(\ell)}) - 1 \\
= & \sum_{n=0}^{\ell-1} \sum_{m=0}^{n} (-1)^m \binom{K}{n-m} \binom{\ell-(n-m)-1}{m} K \int_0^{1/z-1} A^{K-\ell} \int_1^{1+u} \frac{\ln(v)^n \; \ln(1/z)^{-n-1}}{v^2} \, dv \; du \\
- & \sum_{n=0}^{\ell-1} \sum_{m=0}^{n} (-1)^m \binom{K}{n-m} \binom{\ell-(n-m)-1}{m} (n-m) \int_0^{1/z-1} A^{K-\ell} \int_1^{1+u} \frac{\ln(v)^{n-1} \; \ln(1/z)^{-n}}{v^2} \, dv \; du \\
= & \sum_{m=0}^{\ell-1} (-1)^m \binom{K}{\ell-1-m} \binom{\ell-(\ell-1-m)-1}{m} K \int_0^{1/z-1} A^{K-\ell} \int_1^{1+u} \frac{\ln(v)^{\ell-1} \; \ln(1/z)^{-\ell}}{v^2} \, dv \; du \\
+ & \sum_{n=0}^{\ell-2} \sum_{m=0}^{n} (-1)^m \binom{K}{n-m} \binom{\ell-(n-m)-1}{m} K \int_0^{1/z-1} A^{K-\ell} \int_1^{1+u} \frac{\ln(v)^n \; \ln(1/z)^{-n-1}}{v^2} \, dv \; du \\
- & \sum_{n=1}^{\ell-1} \sum_{m=0}^{n-1} (-1)^m \binom{K}{n-m} \binom{\ell-(n-m)-1}{m} (n-m) \int_0^{1/z-1} A^{K-\ell} \int_1^{1+u} \frac{\ln(v)^{n-1} \; \ln(1/z)^{-n}}{v^2} \, dv \; du
\end{aligned}
$$

$$= \sum_{m=0}^{\ell-1} (-1)^m \binom{K}{\ell-1-m} \binom{\ell-(\ell-1-m)-1}{m} K \int_0^{1/z-1} A^{K-\ell} \int_1^{1+u} \frac{\ln(v)^{\ell-1} \ln(1/z)^{-\ell}}{v^2} \, dv \, du$$
$$+ \sum_{n=0}^{\ell-2} \sum_{m=0}^{n} (-1)^m \binom{K}{n-m} \binom{\ell-(n-m)-1}{m} K \int_0^{1/z-1} A^{K-\ell} \int_1^{1+u} \frac{\ln(v)^n \ln(1/z)^{-n-1}}{v^2} \, dv \, du$$
$$- \sum_{n=0}^{\ell-2} \sum_{m=0}^{n} (-1)^m \binom{K}{n+1-m} \binom{\ell-(n+1-m)-1}{m} (n+1-m) \int_0^{1/z-1} A^{K-\ell} \int_1^{1+u} \frac{\ln(v)^n \ln(1/z)^{-n-1}}{v^2} \, dv \, du$$

With propositions 2.5 and 2.6 the result follows.

□

Now we are in the position to prove theorem 2.1.
Proof: [of Theorem 2.1] We start off with

$$\begin{aligned} & E(X_{(\ell+1)}/X_{(\ell)}) - 1 \\ = \;& \frac{1}{(\ell-1)!} \prod_{j=0}^{\ell-1} (K-j) \int_0^{1/z-1} \left(1 - \frac{\ln(u+1)}{\ln(1/z)}\right)^{K-\ell} \int_1^{1+u} \frac{\ln(v)^{\ell-1} \ln(1/z)^{-\ell}}{v^2} \, dv \, du \\ = \;& \frac{\ln(1/z)^{-\ell}}{(\ell-1)!} \prod_{j=0}^{\ell-1} (K-j) \int_0^{1/z-1} \left(1 - \frac{\ln(u+1)}{\ln(1/z)}\right)^{K-\ell} \int_1^{1+u} \frac{\ln(v)^{\ell-1}}{v^2} \, dv \, du \end{aligned}$$

If we focus on the inner integral, we find ($\nu = \ln(v)$, $d\nu = \frac{1}{v} dv$, $v = e^{\nu}$)

$$\int_1^{1+u} \frac{\ln(v)^{\ell-1}}{v^2} \, dv = \int_0^{\ln(1+u)} \nu^{\ell-1} \, e^{-\nu} \, d\nu = \gamma(\ell, \ln(1+u))$$

where $\gamma(n,x)$ denotes the (lower) incomplete Γ function. In particular, $\gamma(n,x) = (n-1)\,\gamma(n-1,x) - x^{n-1} e^{-x}$. Thus, for $\ell \geq 2$, using partial integration in the 4'the equality,

$$\begin{aligned} & E(X_{(\ell+1)}/X_{(\ell)}) - 1 \\ = \;& \frac{\ln(1/z)^{-\ell}}{(\ell-1)!} \prod_{j=0}^{\ell-1} (K-j) \int_0^{1/z-1} \left(1 - \frac{\ln(u+1)}{\ln(1/z)}\right)^{K-\ell} \int_1^{1+u} \frac{\ln(v)^{\ell-1}}{v^2} \, dv \, du \\ = \;& \frac{\ln(1/z)^{-\ell}}{(\ell-1)!} \prod_{j=0}^{\ell-1} (K-j) \int_0^{1/z-1} \gamma(\ell, \ln(1+u)) \left(1 - \frac{\ln(u+1)}{\ln(1/z)}\right)^{K-\ell} du \\ = \;& -\frac{\ln(1/z)^{-\ell+1}}{(\ell-1)!} \prod_{j=0}^{\ell-2} (K-j) \int_0^{1/z-1} \gamma(\ell, \ln(1+u)) \, (1+u) \frac{d}{du} \left(1 - \frac{\ln(u+1)}{\ln(1/z)}\right)^{K-(\ell-1)} du \\ = \;& \frac{\ln(1/z)^{-\ell+1}}{(\ell-1)!} \prod_{j=0}^{\ell-2} (K-j) \int_0^{1/z-1} \left(\gamma(\ell, \ln(1+u)) + \frac{\ln(1+u)^{\ell-1}(1+u)}{(1+u)^2}\right) \left(1 - \frac{\ln(u+1)}{\ln(1/z)}\right)^{K-(\ell-1)} du \\ = \;& \frac{\ln(1/z)^{-\ell+1}}{(\ell-1)!} \prod_{j=0}^{\ell-2} (K-j) \int_0^{1/z-1} \Big((\ell-1)\gamma(\ell-1, \ln(1+u))\Big) \left(1 - \frac{\ln(u+1)}{\ln(1/z)}\right)^{K-(\ell-1)} du \\ = \;& \frac{\ln(1/z)^{-(\ell-1)}}{(\ell-2)!} \prod_{j=0}^{\ell-2} (K-j) \int_0^{1/z-1} \gamma(\ell-1, \ln(1+u)) \left(1 - \frac{\ln(u+1)}{\ln(1/z)}\right)^{K-(\ell-1)} du \\ = \;& E(X_{(\ell)}/X_{(\ell-1)}) - 1. \end{aligned}$$

Per finite induction we find that $E(X_{(\ell)}/X_{(\ell-1)})$ is independent off ℓ (for those ℓ that are feasible). If we take $\ell = 1$, we have ($y = z(1+u)$, $\theta = K/\ln(1/z)$)

$$\begin{aligned}
E(X_{(2)}/X_{(1)}) &= 1 + \frac{K}{\ln(1/z)} \int_0^{1/z-1} \left(1 - \frac{\ln(u+1)}{\ln(1/z)}\right)^{K-1} \int_1^{1+u} \frac{1}{v^2}\; dv\; du \\
&= 1 + \frac{K}{\ln(1/z)} \int_0^{1/z-1} \left(\frac{1}{z} - \frac{1}{z(1+u)}\right) \left(\frac{\ln(1/(z(1+u)))}{\ln(1/z)}\right)^{K-1} z\, du \\
&= 1 + \theta \int_z^1 \left(\frac{1}{z} - \frac{1}{y}\right) \left(\frac{\ln(1/y)}{\ln(1/z)}\right)^{\theta \ln(1/z)-1} dy = G(\theta, z).
\end{aligned}$$

□

2.3.3 Limit $z \to 0$

We expect this expression mainly to depend on $\theta = K/\ln(1/z)$. That is, our formula reads

$$E(X_{(\ell+1)}/X_{(\ell)}) = G(\theta, z) = 1 + \theta \int_z^1 \left(\frac{1}{z} - \frac{1}{y}\right) \left(\frac{\ln(1/y)}{\ln(1/z)}\right)^{\ln(1/z)\,\theta-1} dy$$

In order to discuss the dependencies of $G(\theta, z)$, we keep θ fixed and take the limit $z \to 0$.

Proposition 2.8

$$\lim_{z\to 0} G(\theta, z) = \frac{\theta}{\theta - 1} \tag{14}$$

Proof: We use the transformation $y = z(1+u)$, and $w = \ln(1+u)$, and introduce $x = \ln(1/z)$:

$$\begin{aligned}
&\lim_{z\to 0} \int_z^1 \left(\frac{1}{z} - \frac{1}{y}\right) \left(\frac{\ln(1/y)}{\ln(1/z)}\right)^{\ln(1/z)\,\theta-1} dy \\
=\; &\lim_{z\to 0} \int_0^{1/z-1} \left(\frac{1}{z} - \frac{1}{z(1+u)}\right) \left(\frac{\ln(1/(z(1+u)))}{\ln(1/z)}\right)^{\ln(1/z)\,\theta-1} z\, du \\
=\; &\lim_{z\to 0} \int_0^{1/z-1} \frac{u}{1+u} \left(1 - \frac{\ln(1+u)}{\ln(1/z)}\right)^{\ln(1/z)\,\theta-1} du \\
=\; &\lim_{x\to\infty} \int_0^x (e^w - 1) \left(1 - \frac{w}{x}\right)^{x\,\theta-1} dw
\end{aligned}$$

In order to compute this limit, we first note that $\zeta(u) = \ln(1-u) + u/(1-u)$ has the derivative $\zeta'(u) = u/(1-u)^2 \geq 0$ for $u \in [0,1)$; since $\zeta(0) = 0$, we have $\zeta(u) \geq 0$ for $u \in [0,1)$. Since $\frac{d}{dx}(1-w/x)^x = (1-w/x)^x\, \zeta(w/x) \geq 0$, we have $0 \leq (1-w/x)^x \leq e^{-w}$, and $(1-w/x)^x$ tends in a monotonously increasing way to e^{-w}.

Next we note (recall that $\theta > 1$)

$$\limsup_{x\to\infty} \int_0^x (e^w - 1) \left(1 - \frac{w}{x}\right)^{x\,\theta} dw \leq \limsup_{x\to\infty} \int_0^x (e^w - 1)\; e^{-\theta\, w}\, dw = \int_0^\infty (e^w - 1)\; e^{-\theta\, w}\, dw < \infty.$$

Furthermore, for any $x_0 \in \mathbb{R}$, we find

$$\liminf_{x\to\infty} \int_0^x (e^w - 1) \left(1 - \frac{w}{x}\right)^{x\,\theta} dw \geq \liminf_{x\to\infty} \int_0^{x_0} (e^w - 1) \left(1 - \frac{w}{x}\right)^{x\,\theta} dw = \int_0^{x_0} (e^w - 1)\; e^{-\theta w}\, dw.$$

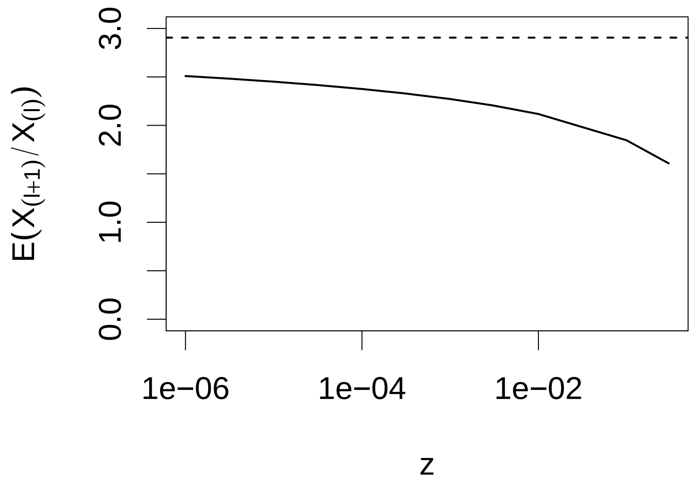


Figure 2: Convergence of $G(\theta, z)$ (solid line) to $G(\theta, 0)$ (dashed line) for $\theta = 1.5$. Note the logarithmic scale of the x-axis.

As $\int_0^{x_0} (e^w - 1)\; e^{-\theta w}\, dw \quad \to \quad \int_0^{\infty} (e^w - 1)\; e^{-\theta w}\, dw$ for $x_0 \quad \to \quad \infty$, we conclude that $\int_0^{x} (e^w - 1)\; \left(1 - \frac{w}{x}\right)^{x\,\theta} dw$ converges to $\int_0^{\infty} (e^w - 1)\; e^{-\theta w}\, dw$.
Last, we consider the limiting behaviour of

$$\int_0^x (e^w - 1)\; \left(1 - \frac{w}{x}\right)^{x\,\theta}\; \left(1 - \frac{w}{x}\right)^{-1} dw = \int_0^x (e^w - 1)\; \left(1 - \frac{w}{x}\right)^{x\,(\theta - \varepsilon)}\; \left(1 - \frac{w}{x}\right)^{\varepsilon\, x - 1} dw.$$

Let $\varepsilon > 0$, s.t. $\theta - \varepsilon > 1$. For x large enough, $\varepsilon\, x - 1 > 0$ and $\left(1 - \frac{w}{x}\right)^{\varepsilon\, x - 1} < 1$. With the argument from above,

$$\limsup_{x \to \infty} \int_0^x (e^w - 1)\; \left(1 - \frac{w}{x}\right)^{x\,\theta}\; \left(1 - \frac{w}{x}\right)^{-1} dw \le \int_0^\infty (e^w - 1)\; e^{-(\theta - \varepsilon) w}\, dw.$$

Since this inequality holds true for any $\varepsilon > 0$, we can take $\varepsilon = 0$. The estimate for lim inf from below relies on the same argument as before: for $x_0 \in \mathbb{R}$ fixed we have

$$\begin{aligned} &\liminf_{x \to \infty} \int_0^x (e^w - 1)\; \left(1 - \frac{w}{x}\right)^{x\,\theta}\; \left(1 - \frac{w}{x}\right)^{-1} dw \\ \ge\; &\liminf_{x \to \infty} \int_0^{x_0} (e^w - 1)\; \left(1 - \frac{w}{x}\right)^{x\,\theta}\; \left(1 - \frac{w}{x}\right)^{-1} dw = \int_0^{x_0} (e^w - 1)\; e^{-\theta\, w}\, dw. \end{aligned}$$

Hence,

$$\lim_{x \to \infty} \int_0^x (e^w - 1)\; \left(1 - \frac{w}{x}\right)^{x\,\theta - 1} dw = \int_0^\infty (e^w - 1)\, e^{-\theta w}\, dw = \frac{1}{\theta - 1} - \frac{1}{\theta}.$$

□

For $n \gg 1$ and $z \ll 1$ and $\theta = K/\ln(1/z)$, we have

$$E(X_{(\ell+1)}/X_{(\ell)}) \quad \approx \quad \frac{\theta}{\theta - 1} \tag{15}$$

From this result, we conclude that $E(X_{(\ell+1)}/X_{(\ell)})$ mainly depends on $\theta = K/\ln(1/z)$; however, the convergence of $G(\theta, z)$ to $G(\theta, 0)$ is rather slow (see figure 2). If we parametrize the model with the rescaled "party creation rate" θ and the relative minimal party size z, the model is sensitive in θ and insensitive in z. It is only necessary to know the rough magnitude of z as long as we know θ precisely.

2.3.4 Logarithm of the group sizes

Proof: [of theorem 2.3] Using proposition 2.4, we find

$$\lim_{n\to\infty} E(\ln(X_{(1)}) - \ln(n)) = \lim_{n\to\infty} \sum_{i=n_0}^{n} \ln(i/n) P(X_{(1)} = i)$$

$$= \lim_{n\to\infty} \sum_{i=n_0}^{n} \ln(i/n) \left\{ \left(\sum_{j=i}^{n} \frac{c}{j} \right)^K - \left(\sum_{j=i+1}^{n} \frac{c}{j} \right)^K \right\} = \lim_{n\to\infty} \sum_{i=n_0}^{n} \ln(i/n) \left\{ \left(\sum_{j=i}^{n} \frac{c}{j} \right)^K - \left(\sum_{j=i}^{n} \frac{c}{j} - \frac{c}{i} \right)^K \right\}.$$

Note that $\sum_{j=i}^{n} \frac{c}{j} = \mathcal{O}(n^0)$. Taylor expansion yields

$$\begin{aligned} \lim_{n\to\infty} E(\ln(X_{(1)}) - \ln(n)) &= \lim_{n\to\infty} \sum_{i=n_0}^{n} \ln(i/n) \left\{ \frac{K\,c}{i} \left(\sum_{j=i}^{n} \frac{c}{j} \right)^{K-1} + \mathcal{O}(i^{-2}) \right\} \\ &= \lim_{n\to\infty} \sum_{i=n_0}^{n} \ln(i/n) \left\{ \frac{K\,c^K}{i/n} \left(\sum_{j=i}^{n} \frac{1}{j/n} \frac{1}{n} \right)^{K-1} + \mathcal{O}((i/n)^{-2}) \frac{1}{n} \right\} \frac{1}{n} \end{aligned}$$

We recognize two nested Riemann sums, that converge to the corresponding integrals. If we use that c converges to $1/\ln(1/z)$ for $n \to \infty$, we obtain

$$\lim_{n\to\infty} E(\ln(X_{(1)}) - \ln(n)) \quad = \quad K \ln(1/z)^{-K} \int_z^1 \frac{\ln(x)}{x} \left(\int_x^1 \frac{1}{y}\, dy \right)^{K-1} dx = -K \int_z^1 \frac{1}{x} \left(\frac{\ln(1/x)}{\ln(1/z)} \right)^K dx.$$

For $\ell > 1$, $E(\ln(X_\ell))$ is handled in a similar way:

$$\lim_{n\to\infty} E(\ln(X_{(\ell)}) - \ln(n)) = \sum_{i=n_0}^{n} \ln(i/n) P(X_{(\ell)} = i)$$

$$= \lim_{n\to\infty} \sum_{i=n_0}^{n} \left\{ \ln(i/n) \frac{c^K}{i} \sum_{j=0}^{\ell-1} \binom{K}{j} \left\{ \left(\sum_{w=n_0}^{i-1} \frac{1}{w} \right)^{j-1} \left(\sum_{m=i+1}^{n} \frac{1}{m} \right)^{K-j-1} \left[K \left(\sum_{w=n_0}^{i-1} \frac{1}{w} \right) - jc^{-1} \right] \right\} + \mathcal{O}(i^{-2}) \right\}$$

$$= -\sum_{j=0}^{\ell-1} \binom{K}{j} \int_z^1 \frac{1}{x} \left(1 - \frac{\ln(1/x)}{\ln(1/z)} \right)^{j-1} \left(\frac{\ln(1/x)}{\ln(1/z)} \right)^{K-j} \left(K - j - K \frac{\ln(1/x)}{\ln(1/z)} \right) dx.$$

□

Note that $E(\log(X_{(\ell+1)})) - E(\log(X_{(\ell)}))$ is not independent on ℓ, even for n large; the expectation of $\log(X_{(\ell)})$ does not depend exactly in a linear way on the rank ℓ, but only approximately. The growth law of the group sizes can be better seen in the rations of subsequent groups than in the logarithm of group sizes. However, for practical purpose, the difference of the linear growth law for the logarithmic group sizes is negligible.

As as heuristic estimator for z we will use

$$E(\ln(X_{(1)}) - \ln(n) = -K \int_z^1 \frac{1}{x}\left(\frac{\ln(1/x)}{\ln(1/z)}\right)^K dx.$$

We replace $E(X_{(1)})$ by the minimal observed group size (to be more precise: by the minimal group size predicted by a linear fit of the data), and infer from the relation above the parameter z. Though we consider data that are conditioned on the total population (number of voters known), for practical purposes this estimator works fine (see figures in section 7).

2.4 Rank statistics – conditioned case

Now we investigate the corresponding order statistics in the conditioned case: We consider K independent realizations $X_1, \dots, X_K$ of i.i.d. RV that assume values in $\{n_0, \dots, n\}$, where $0 < n_0 < n$, $P(X_i = j) = c/j$ for $j \in \{n_0, \dots, n\}$ and 0 else, where $c^{-1} = \sum_{j=n_0}^n j^{-1}$. We condition on $\sum_{i=1}^K X_i = n$ and order these realizations according to size $X_{(1)} \leq X_{(2)} \dots \leq X_{(K)}$. In order to distinguish the conditioned and the non-conditioned random variables, let us denote the realizations with condition by

$$X_{(1),n} \leq X_{(2),n} \dots \leq X_{(K),n}.$$

The objects to investigate are $E(X_{(\ell+1),n}/X_{(\ell),n})$ and $E(\ln(X_{(\ell),n}))$.

2.4.1 Joint distribution

Proposition 2.9 *Let*

$$M_K = \Bigg\{(i_1, \dots i_K)\,|\, n_0 \leq i_1 \leq n/K, \quad i_K = n - \sum_{\ell=1}^K i_\ell, \qquad i_{j-1} \leq i_j \leq \frac{1}{K-j+1}\left(n - \sum_{\ell=1}^{j-1} i_\ell\right), \quad j = 2, \dots K-1\Bigg\}$$

and

$$c_K \approx \left(\int_z^{1/K}\int_{x_1}^{(1-x_1)/(K-1)}\int_{x_2}^{(1-x_1-x_2)/(K-2)} \cdots \int_{x_{K-1}}^{(1-\sum_{j=1}^{K-2} x_j)/2} \frac{1}{1-\sum_{\ell=1}^{K-1} x_\ell}\prod_{\ell=1}^{K-1}\frac{1}{x_\ell}\, dx_{K-1}\cdots dx_1\right)^{-1}.$$

Then, for $(i_1, \dots, i_K) \in M_K$, we have

$$P(X_{(1),n} = i_1, \dots, X_{(K),n} = i_K) = c_K \prod_{\ell=1}^K \frac{1}{i_\ell} + \mathcal{O}(n^{-1}).$$

Proof: The values that $(X_{(1),n} \leq X_{(2),n} \dots \leq X_{(K),n})$ can assume is given by $(i_1, \dots i_K) \in M_K$ with

$$M_k = \{(i_1, \dots i_K)\,|\, n_0 \leq i_1 \leq i_2 \dots \leq i_K, \sum_{\ell=1}^K i_\ell = n\}.$$

In order to obtain a conditioned realization, we may draw unconditioned realizations until the condition is hit, and only accept those. Hence, the probability for an admissible value is proportional to the unconditioned probability distribution:

$$P(X_{(1),n} = i_1, \ldots X_{(K),n} = i_K) \;\; = \;\; C \prod_{\ell=1}^{K} \frac{1}{i_\ell};$$

The constant C can be determined by

$$C^{-1} = \sum_{(i_1,\ldots i_K)\in M_K} \prod_{\ell=1}^{K} \frac{1}{i_\ell}.$$

We characterize M_K better. For $(i_1, \ldots i_K) \in M_K$, we mey write

$$i_K = n - \sum_{\ell=1}^{K-1} i_\ell.$$

Thus,

$$i_{K-1} \le i_K = n - \sum_{\ell=1}^{K-1} i_\ell \quad \Rightarrow \quad i_{K-1} \le \frac{1}{2}\left(n - \sum_{\ell=1}^{K-2} i_\ell\right).$$

We can proceed recursively,

$$\begin{aligned} i_{K-2} \le i_{K-1} \;\; &\le \;\; \frac{1}{2}\left(n - \sum_{\ell=1}^{K-2} i_\ell\right) \quad \Rightarrow \quad i_{K-2} \le i_{K-1} \le \frac{1}{3}\left(n - \sum_{\ell=1}^{K-3} i_\ell\right) \\ \cdots \quad i_{K-j-1} \le i_{K-j} \;\; &\le \;\; \frac{1}{j+1}\left(n - \sum_{\ell=1}^{K-j-1} i_\ell\right) \end{aligned}$$

or

$$i_{j-1} \le i_j \le \frac{1}{K-j+1}\left(n - \sum_{\ell=1}^{j-1} i_\ell\right)$$

For i_1, we obtain the maximal value given if all indices are equal, $i_1 \le n/K$. Hence,

$$\begin{aligned} M_K \;\; = \;\; \Bigg\{ &(i_1, \ldots i_K) \,|\, n_0 \le i_1 \le n/K, \quad i_K = n - \sum_{\ell=1}^{K} i_\ell, \\ &\qquad\qquad i_{j-1} \le i_j \le \frac{1}{K-j+1}\left(n - \sum_{\ell=1}^{j-1} i_\ell\right), \quad j = 2, \ldots K-1 \Bigg\}. \end{aligned}$$

Therewith,

$$\begin{aligned} C^{-1} &= \sum_{(i_1,\ldots i_K)\in M_K} \frac{1}{1 - \sum_{\ell=1}^{K-1} i_\ell/n} \prod_{\ell=1}^{K-1} \frac{1}{i_\ell/n}\, n^{-K} \\ &\approx \;\; n^{-1} \int_z^{1/K} \int_{x_1}^{(1-x_1)/(K-1)} \int_{x_2}^{(1-x_1-x_2)/(K-2)} \cdots \int_{x_{K-1}}^{(1-\sum_{j=1}^{K-2} x_j)/2} \frac{1}{1 - \sum_{\ell=1}^{K-1} x_\ell} \prod_{\ell=1}^{K-1} \frac{1}{x_\ell}\, dx_{K-1} \cdots dx_1 \\ &=: \;\; (c_K\, n)^{-1}. \end{aligned}$$

□

In order to compute the constant c_K, we can waive the order of x_i. As there are $K!$ possibilities to order a vector $(x_1, \ldots, x_K) \in [z_0, 1-(K-1)z_0]^K$, for symmetry reasons, we may write as well

$$c_K^{-1} \;=\; \frac{1}{K!} \int_{z_0}^{1-(K-1)z_0} \cdots \int_{z_0}^{1-(K-1)z_0} \frac{1}{1-\sum_{\ell=1}^{K-1} x_\ell} \prod_{\ell=1}^{K-1} \frac{1}{x_\ell}\, dx_{x_{K-1}} \cdots dx_1.$$

Note that $(X_1, .., X_K)/n$ follows for $n \to \infty$ a truncated Dirichlet distribution with parameters $\alpha_i = 0$; while for the original Dirichlet distribution necessarily $\alpha_i > 0$ due to integrability conditions, the truncated Dirichet distribution is also well defined for $\alpha_i = 0$.

2.4.2 Size ratio

We do not compute the expectation of the quotient for general K but only for $K = 3$. As discussed above, the joint distribution of $(X_{(1)}, X_{(2)})$ is given by

$$P(X_{(1)} = i_1, X_{(2)} = i_2) = \frac{c_3}{i_1\, i_2 (n - i_1 - i_2)}$$

respectively

$$P(X_{(2)} = i_2, X_{(3)} = i_3) = \frac{c_3}{(n - i_2 - i_3)\, i_2\, i_3}$$

where

$$c_3^{-1} = \frac{2}{3!} \int_z^{1-2z} \frac{1}{x(1-x)}\, \ln\left(\frac{1-z-x}{z}\right) dx.$$

Therewith,

$$c_3 \lim_{n\to\infty} E(X_{(2)}/X_{(1)}) \;=\; \int_{z_0}^{1/3} \int_x^{(1-x)/2} \frac{1}{x^2\,(1-x-y)}\, dy\, dx = \int_{z_0}^{1/3} x^{-2}\; \ln\left(\frac{2(1-2x)}{(1-x)}\right) dx.$$

Furthermore,

$$c_3 \lim_{n\to\infty} E(X_{(3)}/X_{(2)}) \;=\; \int_{z_0}^{1/3} \int_x^{(1-x)/2} \frac{1}{x\; y^2}\, dy\, dx = 2\ln(2) - 1 + \frac{1}{z_0} - 2\ln(1/z_0 - 1).$$

Obviously,

$$\lim_{n\to\infty} E(X_{(2)}/X_{(1)}) \neq \lim_{n\to\infty} E(X_{(3)}/X_{(2)}).$$

The magnitude of both expectation are in the same range, though (see figure 3).

2.4.3 Logarithm of group sizes

Theorem 2.10 *For $k < K$,*

$$\begin{aligned} & E(\ln(X_{(k),n})) - \ln(n) \\ = \;& c_K \int_z^{1/K} \int_{x_1}^{(1-x_1)/(K-1)} \int_{x_2}^{(1-x_1-x_2)/(K-2)} \cdots \int_{x_{K-1}}^{(1-\sum_{j=1}^{K-2} x_j)/2} \frac{\ln(x_k)}{1-\sum_{\ell=1}^{K-1} x_\ell} \prod_{\ell=1}^{K-1} \frac{1}{x_\ell}\, dx_{K-1} \cdots dx_1 \end{aligned}$$

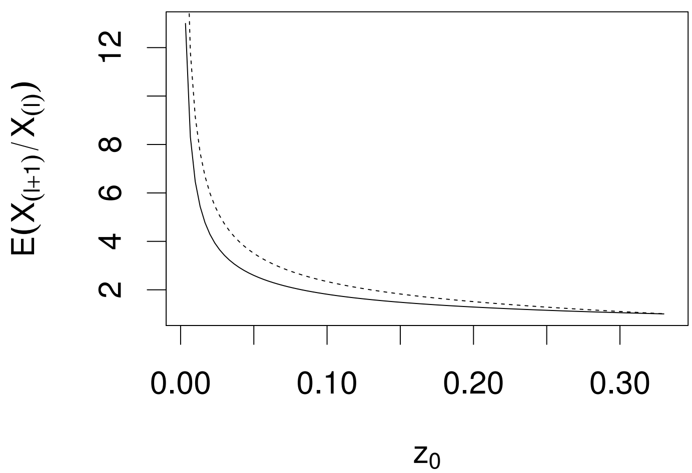


Figure 3: $E(X_{(2)}/X_{(1)})$ (solid line) and $E(X_{(3)}/X_{(2)})$ (dashed line) for $K = 3$.

and

$$
\begin{aligned}
& E(\ln(X_{(K)})) - \ln(n) \\
= \;& c_K \int_z^{1/K} \int_{x_1}^{(1-x_1)/(K-1)} \int_{x_2}^{(1-x_1-x_2)/(K-2)} \cdots \int_{x_{K-1}}^{(1-\sum_{j=1}^{K-2} x_j)/2} \frac{\ln\left(1 - \sum_{\ell=1}^{K-1} x_\ell\right)}{1 - \sum_{\ell=1}^{K-1} x_\ell} \prod_{\ell=1}^{K-1} \frac{1}{x_\ell} \, dx_{K-1} \cdots dx_1.
\end{aligned}
$$

The proof consists of an obvious calculation. Particularly, for n large, $E(\ln(X_{(\ell),n})) = \ln(n)$ plus a term only depending on z, ℓ, K, but not on n.

Though we see that, strictly spoken, there is no linear relation between the logarithmic size of groups and their order, simulations indicate that the dependency is almost linear, even if K is small (see figure 4).

2.4.4 Simulations

Direct simulations of $X_{(k),n}/n$ in a naive way is costly for large n. The convergence of $X_{(k),n}/n$ to the truncated Dirichlet distribution opens the way for a simple method to construct realizations. We fix a population size of $\hat{n} = 10^{5+m}$, where $m \in \mathbb{N}_0$ is the minimal non-negative integer to ensure that $z\,\hat{n} > 5$. Then, we draw K independent realizations of X_k as introduced in section 2.3. We accept a realization if $\sum_{i=1}^K X_k \in [0.975\,\hat{n}, 1.025\,\hat{n}]$. In order to obtain (approximate) realizations for population size n, we rescale $n\,X_k/\hat{n}$. This algorithm is able to handle the populations sizes at hand for the data we consider.

We know that the expected group size of the model does not exactly follow a linear log-rank relation. We investigate the residuals of the data. In order to ensure that the residuals of the model do not show a distinct trend (s.t. all trends in the deviance observed in the data come from a mechanism not

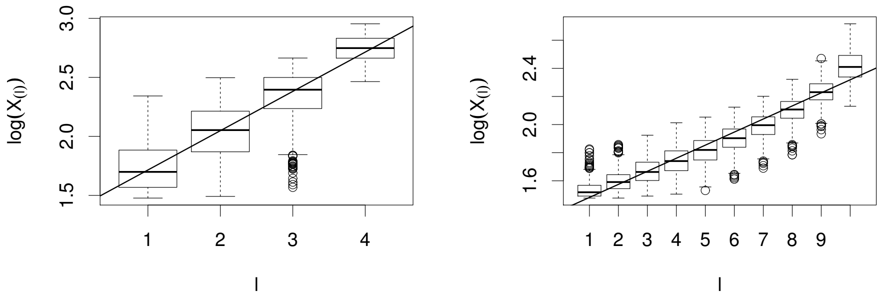


Figure 4: $E(log(X_{(\ell)}))$ over ℓ. Boxplot from 1000 realizations, line is a linear fit to the expectations of $E(log(X_{(\ell)}))$; $n = 1000$, $z = 0.03$. Left:$K = 4$. Right: $K = 10$.

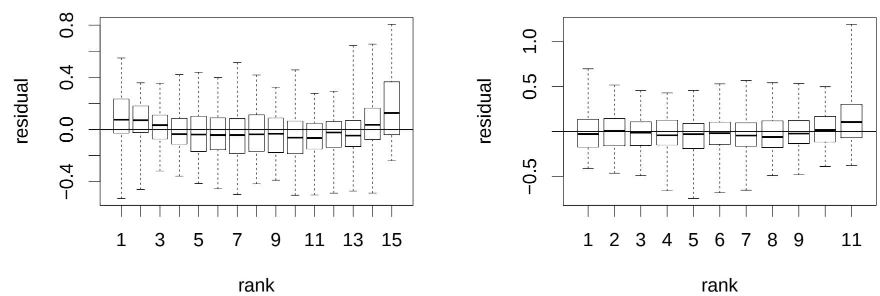


Figure 5: Residuals from the linear fit for 100 runs. In each run, the linear fit and the residuals are determined. Population size used is $n = 10^5$, $z = 0.0004$. (a) Number of parties is fixed, $K = 15$, (b) In each run, K is randomly chosen between 10 and 20. Only the residuals of the 11 largest ranks are shown (largest rank at 11).

covered by the model), we simulate the residuals (Figure 5). In subfigure (a), the number of parties is constant $K = 15$. In order to come closer to the residuals in the data (where the number of parties can be different in each election), in subfigure (b) the number of parties is randomly chosen for each realization. In both cases, the systematic bias is rather small, and does not play a role in comparison with the systematic bias found in the data (Figure 2 in main text).

3 Sensitivity analysis for z

In order to obtain an impression about the influence of the model parameters, in particular the model parameter z (as K and n are given explicitly in the data), we take the election in the Netherlands (2017), and compare that data with model predictions, where we vary z (figure 6). We use these recent data as the Netherlands do not have a threshold as the FRG, and in this satisfy the model requirements rather well. As predicted by the theory, the simulated data are approximately linear in

the rank. If we inspect the lower panels in figure 6, we find that the maximum value is hardly affected by z, but the maximum basically meets the observed data. The slope, however, is affected by z. In that, if we vary z, we find a certain range where the slope and the maximum group size predicted by the model agrees with the data; if z is too low or too high, we have a distinct difference.
This observation is also reflected by the distance of data with the simulated mean group sizes (figure 6, top panel). We find that the distance of the mean from the data has a unique minimum; there is, however, a certain plateau where the distance does not differ strongly from the minimum. Our heuristic estimator (see section 2.3.4) for z does not meet the absolute minimum, but is in the acceptable range.

4 Model comparison: Election model versus linear regression model

There are different model comparison and model selection approaches available. Akaike's Information Criterion (AIC) [2] is a model selection criterion commonly used. The smaller the AIC the better the model. Given data, from a purely statistical aspects, a model is better if the likelihood of the data is higher. Clearly, a model with more parameter should be able to better fit the data. The AIC therefore increases the score linearly in the number of parameters. If k is the number of parameters, and $\mathrm{ll}(x|\theta)$ is the likelihood of the data given parameters, the AIC is given by

$$\mathrm{AIC} = 2\,k - 2\,\ln(\mathrm{ll}(x|\theta)).$$

The problem in our case is mainly that the likelihood is not available analytically. We need to determine the likelihood numerically.

4.1 Approximate Likelihood

Assume we have an $\mathbb{R}^K$-valued random variable (with C^1 density φ). We can simulate this random variable (draw realizations), but do not know the density φ explicitly. We have data $\hat{x} \in \mathbb{R}^K$, and would like to know the likelihood of $\hat{x}$, that is, we aim to estimate the density of the random variable at $\hat{x}$.
We draw m samples $x_1, \ldots, x_m$, and choose $\varepsilon > 0$. Next, we compute the fractions of realizations with a distance of at most ε from $\hat{x}$, and find

$$P(\|X - \hat{x}\| < \varepsilon) = \int_{B_\varepsilon(\hat{x})} \varphi(x)\,dx \approx \varphi(\hat{x})\,\omega_K\,\varepsilon^K, \quad P(\|X - \hat{x}\| < \varepsilon) \approx \frac{\#\{x_i\,|\,\|x_i - \hat{x}\| < \varepsilon\}}{m}$$

and thus (ω_K denotes the size of the K-dimensional unit sphere)

$$\varphi(\hat{x}) \approx \frac{\#\{x_i\,|\,\|x_i - \hat{x}\| < \varepsilon\}}{m\,\omega_K\,\varepsilon^K}.$$

A proof that this method converges for $m \to \infty$ and $\varepsilon \to 0$ (in an appropriate manner) can be found in [10]. Unfortunately, for K reasonable large ($K \approx 30$, say) the course of dimension requires n to be very large, such that this approach is not feasible any more. A way out is to focus not on the complete random vector $x \in \mathbb{R}^K$, but select 2 or 3 components, and only consider the marginal distribution of these few components.

4.2 Test of the method with linear regression model

We take the data from the Netherlands 2017 (number of parties have been 28), and do a linear regression,

$$\log 10(X_{(i)}) = a + b\,i + e_i$$

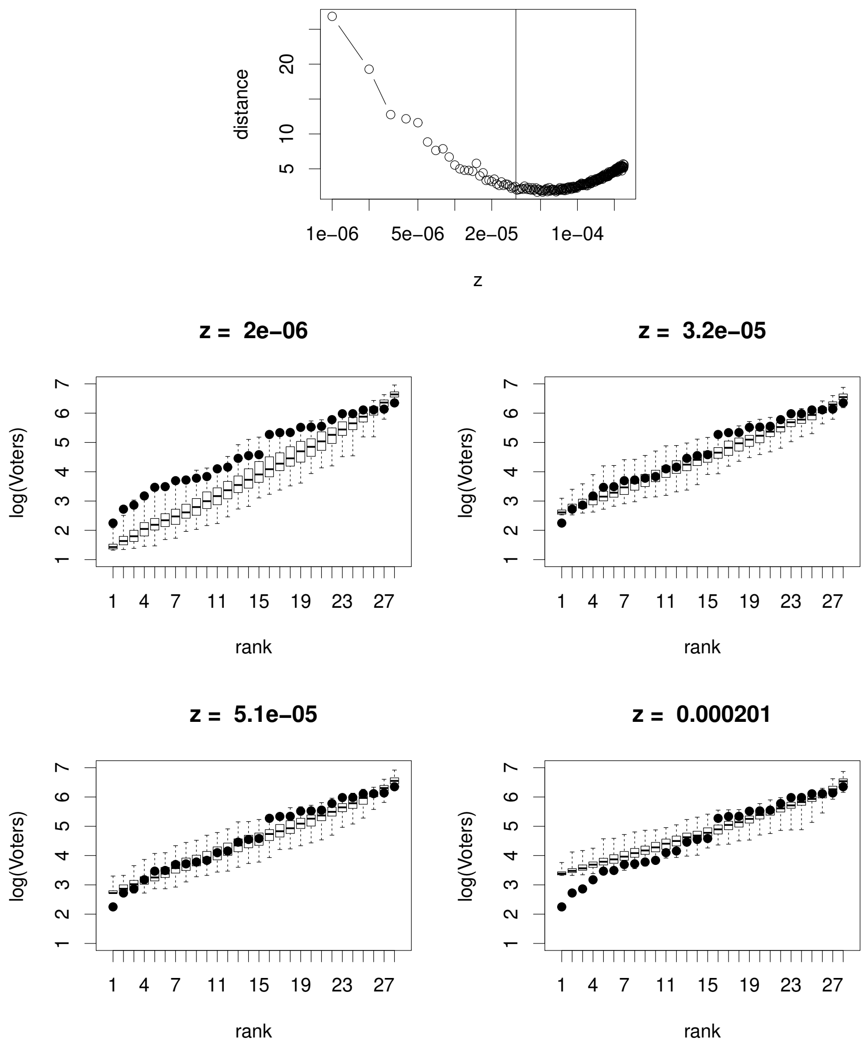


Figure 6: Sensitivity analysis of the model w.r.t. the parameter z. Upper panel: distance of the mean in the simulated data from the data of the Netherlands (2017). Vertical line indicates the location of our heuristic estimator (see section 2.3.4). Panels below: Data (bullets) and boxplot of 100 simulations of the model for indicated values of z.

where e_i are i.i.d., $e_i \sim N(0, \sigma^2)$. We find that

$$\hat{b} = 0.144, \quad \hat{a} = 2.53, \quad \hat{\sigma} = 0.187.$$

If we focus on the marginal likelihood of components 4, 14, and 24 (small medium and large party), we obtain analytically

$$\text{likely}_{\hat{a},\hat{b},\hat{\sigma},(4,14,24)} = 9.18.$$

If we run our algorithm ($\varepsilon = 0.05$, sample size $m = 10^5$), we obtain in three arbitrary runs the approximate likelihood

$$\text{approx.likeli}_{\hat{a},\hat{b},\hat{\sigma},(4,14,24)} : \quad 9.24; \qquad 9.34; \qquad 8.16.$$

The approximation seems to be rather reliable. We obtain the AIC of that model (3 parameters) using the approximate likelihood,

$$AIC_{approx,regression,(4,14,24)} = 1.55; \qquad 1.53; \qquad 1.80,$$

which is consistent with the AIC we find using the analytic likelihood

$$AIC_{regression,(4,14,24)} = 1.57.$$

4.3 Approximate likelihood of the election model and model comparison

In order to draw many realizations from the election model (given the minimal group size z, and the number of parties K resp. number of voters n), we change the strategy to simulate. Above, the aim has been to draw 100 independent realizations. By now, we need to draw thousands of realizations. Therefore, we utilize the fact that the relative groups sizes X_i/n (note that X_i are not ordered) tend for $n \to \infty$ to a truncated K-dimensional Dirichlet distribution with parameters $\alpha_i = 1$ (SI, section 2.4.1). Up to a multiplicative constant, the probability density of that distribution is known. We can use a Metropolis Hastings algorithm [7] to produce (after a burn-in phase which we take to 1000 steps) a time series that is distributed according the truncated Dirichlet distribution. To decrease the autocorrelation of that time series we only store every 10'th realization. These realizations are then multiplied by n, and ordered to address the order structure in our election data.

As before, we consider the data of the election in the Netherlands from 2017. As before we focus on a small (order number 4), medium (order number 14) and large (order number 24) party. For the election model we find the approximate likelihoods (sample size $m = 1e5$, $\varepsilon = 0.1$, 3 runs)

$$\text{approx.likeli}_{election\ model,(4,14,24)} : 1.68; \qquad 2.48; \qquad 2.02$$

corresponding to the AIC

$$AIC_{election\ model,(4,14,24)} = 0.96; \qquad 0.18; \qquad 0.59.$$

In any case, if we compare with $AIC_{regression,(4,14,24)} = 1.57$, we find that the election model is superior. From the purely statistical point of view, this results is based on the numbers of parameters used. The election model has only one parameter that is adapted to the data, while the regression model has three parameters. The likelihood of regression and election model are roughly in the same magnitude.

However, there is a second point, not considered by statistics: the election model is the consequence of a mechanism, while the regression model is a purely phenomenological, generic model. While the mechanistic model provides some hint why we observe the log-linear structure in the election data, the regression model does not address such questions. It is an additional argument in favor of model, that it is based on first principles on a microscopic level (which can be readily discussed) and that these first principles yield on a macroscopic level the observed behavior.

References

[1] R. Arratia, A. Barbour, and S. Tavaré. Poisson process approximations for the Ewens sampling formula. *Ann. Appl. Prob.*, 2:519–535, 1992.

[2] K. P. Burnham, D. R. Anderson, and K. P. Huyvaert. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, 65(1):23–35, aug 2010.

[3] H. Crane. The ubiquitous Ewens sampling formula. *Statistical Science*, 31(1):1–19, feb 2016.

[4] R. Durrett. *Probability Models for DNA Sequence Evolution*. Springer, 2008.

[5] W. Ewens. The sampling theory of selectively neutral alleles. *Theoret. Pop. Biol.*, 3:87–112, 1972.

[6] W. J. Ewens. *Mathematical Population Genetics*. Springer New York, 2004.

[7] W. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, 1996.

[8] A. V. Gnedin. Three sampling formulas. *Combinatorics, Probability and Computing*, 13(2):185–193, mar 2004.

[9] J. F. C. Kingman. Random partitions in population genetics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 361(1704):1–20, may 1978.

[10] F. J. Rubio and A. M. Johansen. A simple approach to maximum intractable likelihood estimation. *Electronic Journal of Statistics*, 7(0):1632–1654, 2013.

[11] O. Z. Simon Tavare. *Lectures on Probability Theory and Statistics*. Springer-Verlag GmbH, 2004.

5 Visualization of Data

5.1 Data sources

The data for the election of the FRG ("Bundestagswahl") can be downloaded from
https://www.bundeswahlleiter.de/en/bundeswahlleiter.html
We only take parties into account that can be vote for by second votes ("Zweitstimmen").

The data for the elections in France can be found in
https://www.interieur.gouv.fr/Elections/Les-resultats

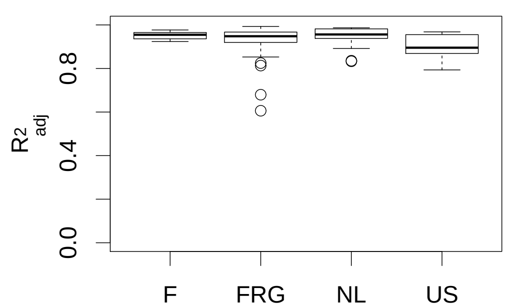
The data for the Republican primaries 2016 in the US can be found in:
Iowa caucuses: https://edition.cnn.com/election/2016/primaries/states/ia
retrieved from
https://www.iowagop.org/
New Hampshire : http://www.thegreenpapers.com/P16/NH-R
retrieved from
http://sos.nh.gov/2016RepPresPrim.aspx?id=8589957185
Nevada: http://www.thegreenpapers.com/P16/NV-R retrieved
from
http://nevadagop.org/nevada-republican-presidential-caucus-results/
Massachusetts: http://www.thegreenpapers.com/P16/MA-R
retrieved from
http://electionstats.state.ma.us/elections/search/year_from:2016/year_to:2016/office_id:1/stage:Republican
Tennessee: http://www.thegreenpapers.com/P16/TN-R
retrieved from
https://sos.tn.gov/products/elections/election-results
Texas: http://www.thegreenpapers.com/P16/TX-R
retrieved from
http://elections.sos.state.tx.us/elchist273_state.htm
Michigan: http://www.thegreenpapers.com/P16/MI-R
retrieved from
http://miboecfr.nictusa.com/election/results/2016PPR_CENR.html
Wisconsin: http://www.thegreenpapers.com/P16/WI-R
retrieved from
http://elections.wi.gov/elections-voting/results/2016/spring-election-presidential-preference

The data for the Netherlands are taken form wikipedida,
https://nl.wikipedia.org/wiki/Tweede_Kamerverkiezingen_1972
https://nl.wikipedia.org/wiki/Tweede_Kamerverkiezingen_1977
https://nl.wikipedia.org/wiki/Tweede_Kamerverkiezingen_1981
https://nl.wikipedia.org/wiki/Tweede_Kamerverkiezingen_1982
https://nl.wikipedia.org/wiki/Tweede_Kamerverkiezingen_1986
https://nl.wikipedia.org/wiki/Tweede_Kamerverkiezingen_1989
https://nl.wikipedia.org/wiki/Tweede_Kamerverkiezingen_1994
https://nl.wikipedia.org/wiki/Tweede_Kamerverkiezingen_1998
https://nl.wikipedia.org/wiki/Tweede_Kamerverkiezingen_2002
https://nl.wikipedia.org/wiki/Tweede_Kamerverkiezingen_2003
https://nl.wikipedia.org/wiki/Tweede_Kamerverkiezingen_2006

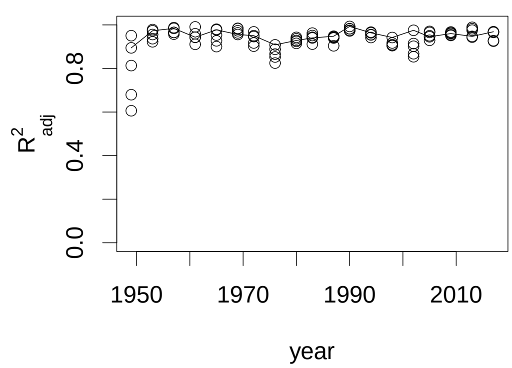
https://de.wikipedia.org/wiki/Parlamentswahl_in_den_Niederlanden_2010
https://de.wikipedia.org/wiki/Parlamentswahl_in_den_Niederlanden_2012
https://de.wikipedia.org/wiki/Parlamentswahl_in_den_Niederlanden_2017

5.2 Overall results

We show below data from the US elections (Republicans, Primaries, 2016, $n = 8$), from France (Presidential elections, first round, 2007, 2012, and 2017, $n = 8$), elections in the Netherlands (1972-2017, n=14), and the Federal Republic of Germany (Federal elections 1949-2017, $n = 95$) from different organizational units (city, federal state, country). We present a semi-logarithmic representation, together with a linear fit of the data. In order to obtain a first, overall impression about the quality of the fits, we consider R^2_{adj} in figure 7. We find that the linear model mostly explains far more than 90% of the variability in the data.



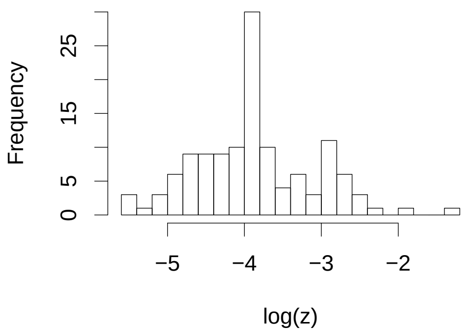
(a)



(b)

Figure 7: (a) Boxplot for the R^2_{adj} values in France ($n = 9$), Germany ($n = 95$), Netherlands ($n = 14$), and the US ($n = 8$). (b) R^2 over year of election for FRG (the values for the complete FRG are connected by a line, the other dots belong to Stuttgart, Munich, Bavaria, and Baden-Württemberg).

In section 7, we fit according to the heuristic estimator (proposed at page 17) the parameter z; the histogram of the logarithm of this parameter obtained from the elections in all four democracies considered is shown in figure 8. We find an unimodal distribution. A closer analysis of the dependency on the number of voters, number of parties for the unit under consideration, and democracy at hand reveals a significant, but weak dependency ($R^2_{adj} = 0.44$; see table in figure 8). Interestingly, also the democracy (France, FRG, the Netherlands and US) has a significant influence on z.



	point estimator	p
intercept	-0.929349	0.078865
voters	-0.204479	0.009044
parties	-0.860429	$4.38e-08$
offset FRG	-0.901228	$3.99e-05$
offset NL	-1.340101	0.000935
offset US	-0.053553	$1.27e-05$

Figure 8: Left: Logarithmic histogram of the parameter z. Right: Result of a linear fit of $\log(z)$ by the number of voters n.

6 Elections in semilogarithmic representation with linear fit

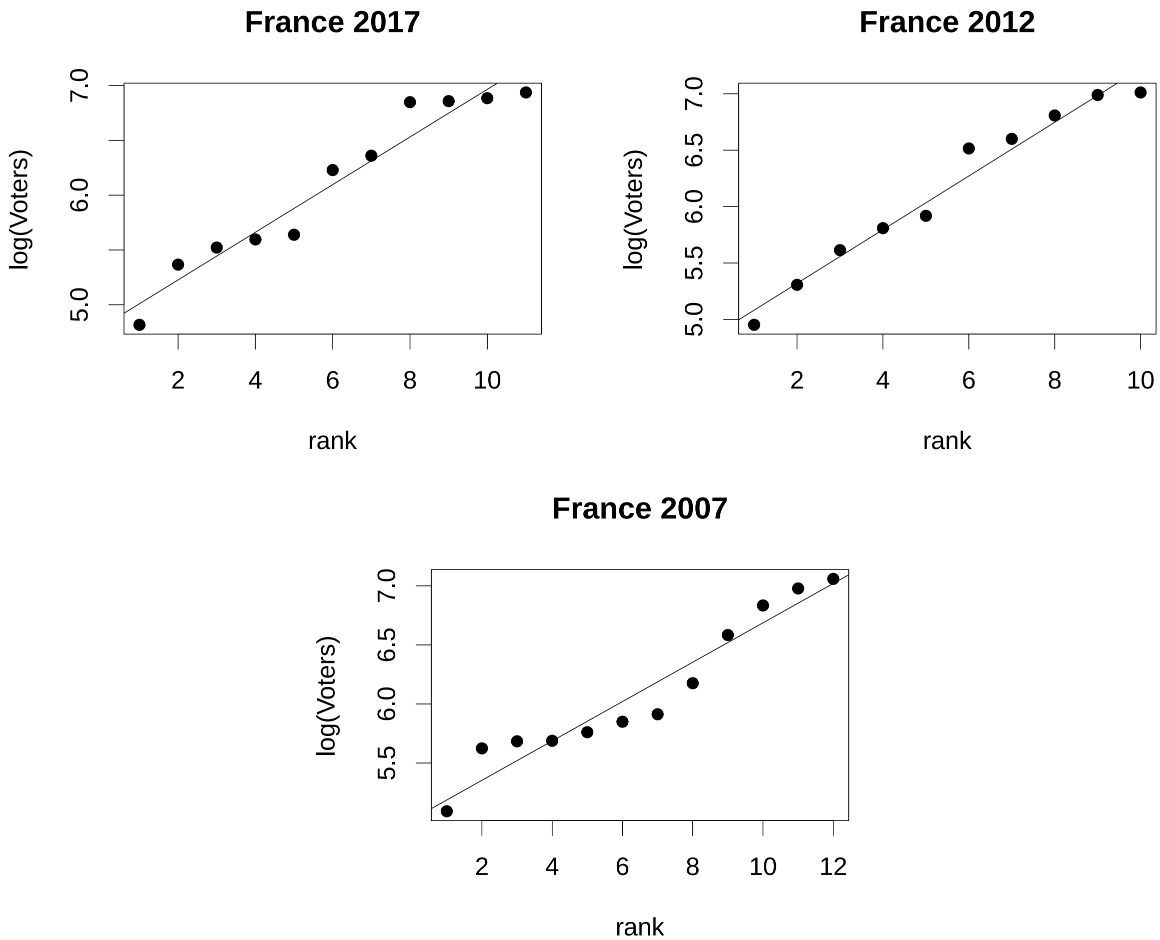


Figure 9: Election France, 2017, 1012, 2007 (bullets: data, line: linear fit).

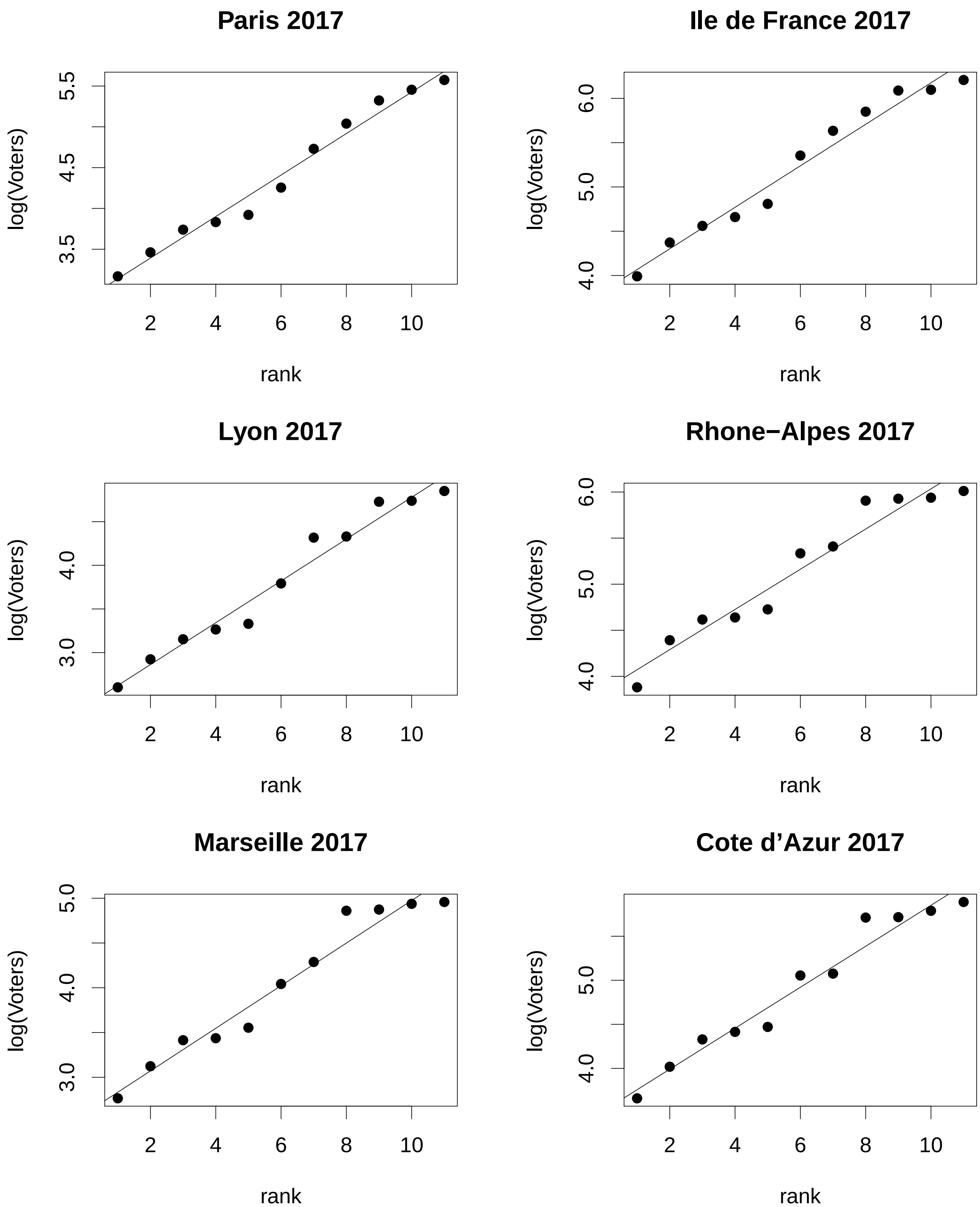


Figure 10: Election France, 2017 (bullets: data, line: linear fit).

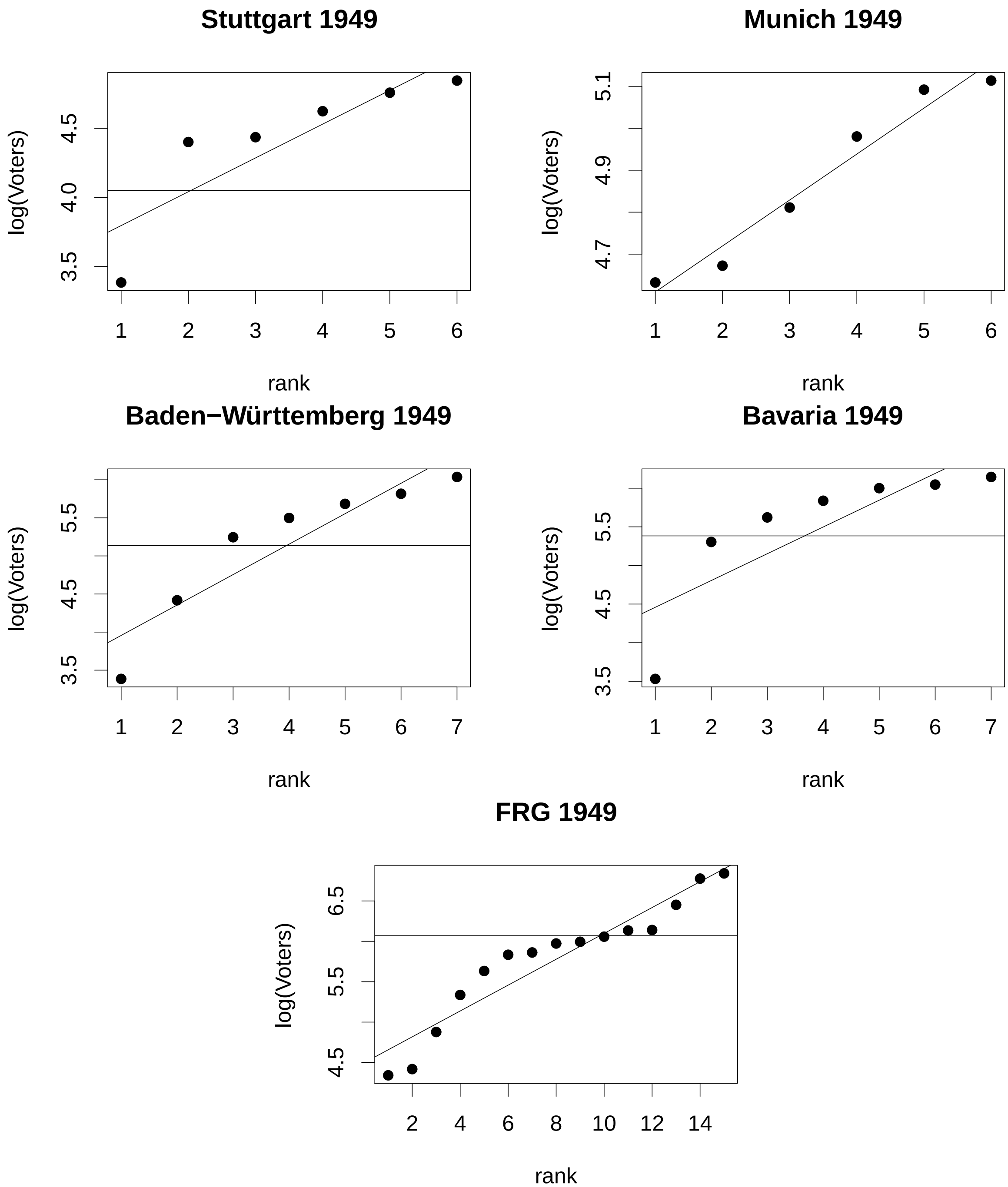


Figure 11: Election FRG, 1949 (bullets: data, line: linear fit).

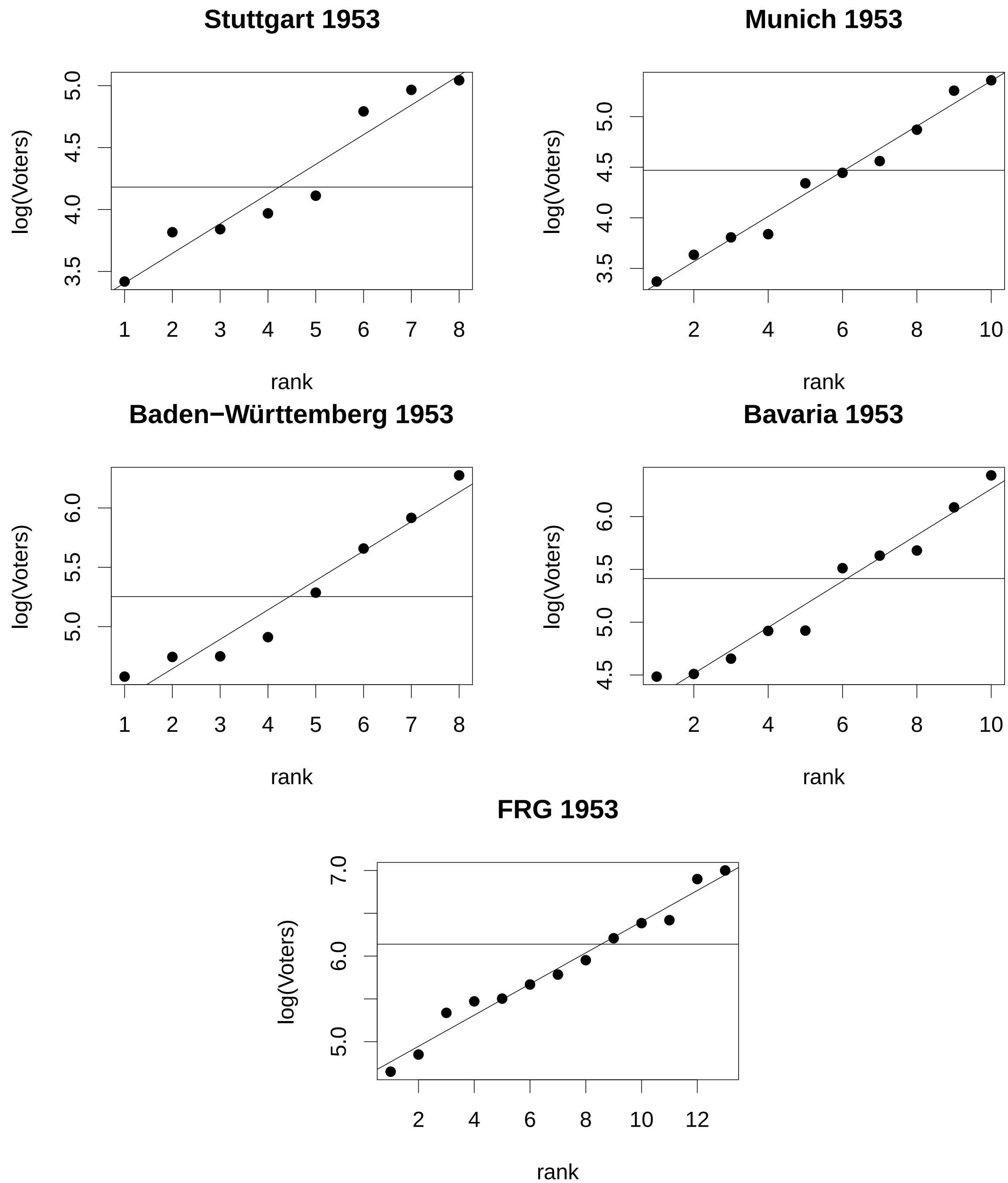


Figure 12: Election FRG, 1953 (bullets: data, line: linear fit).

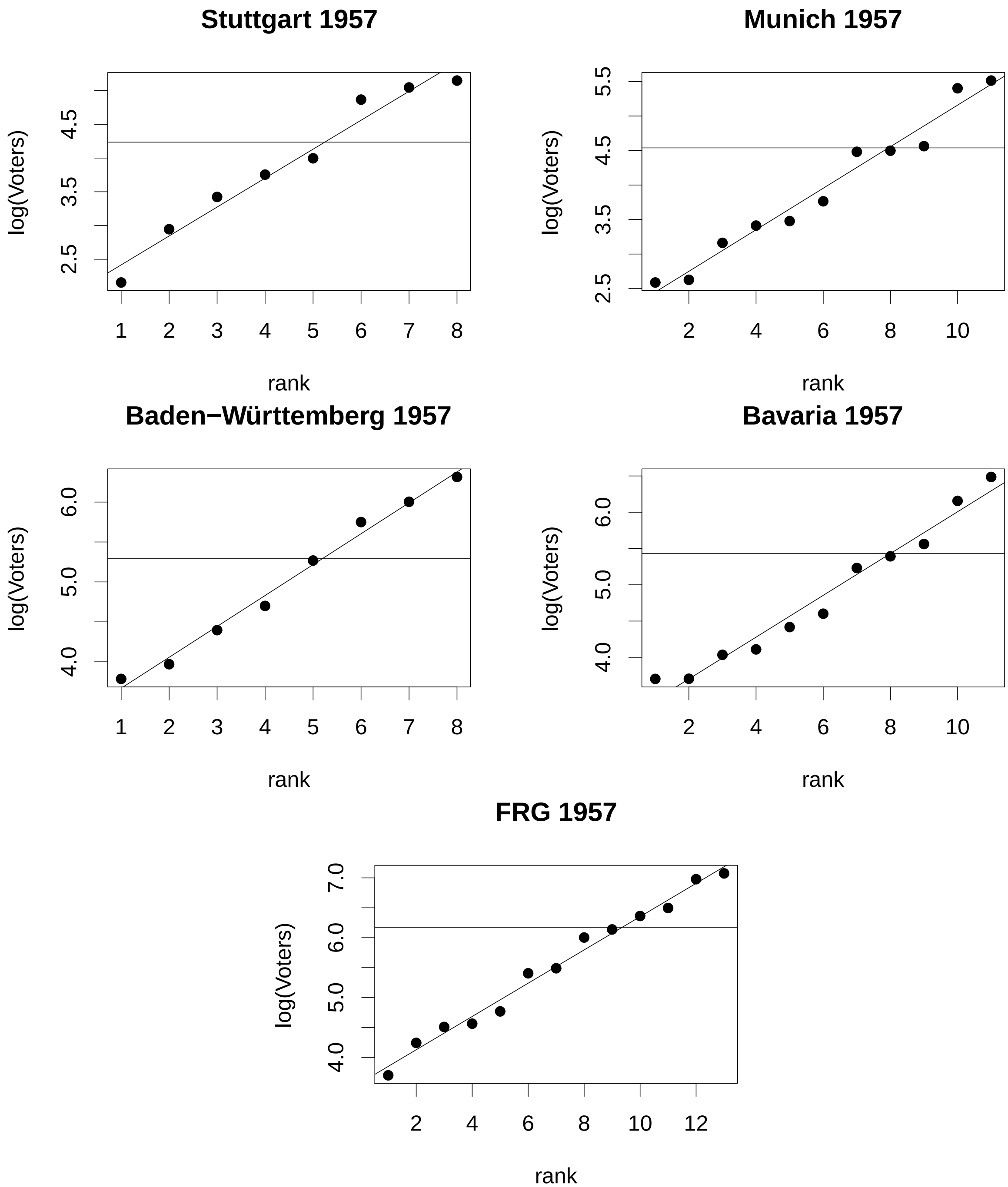


Figure 13: Election FRG, 1957 (bullets: data, line: linear fit).

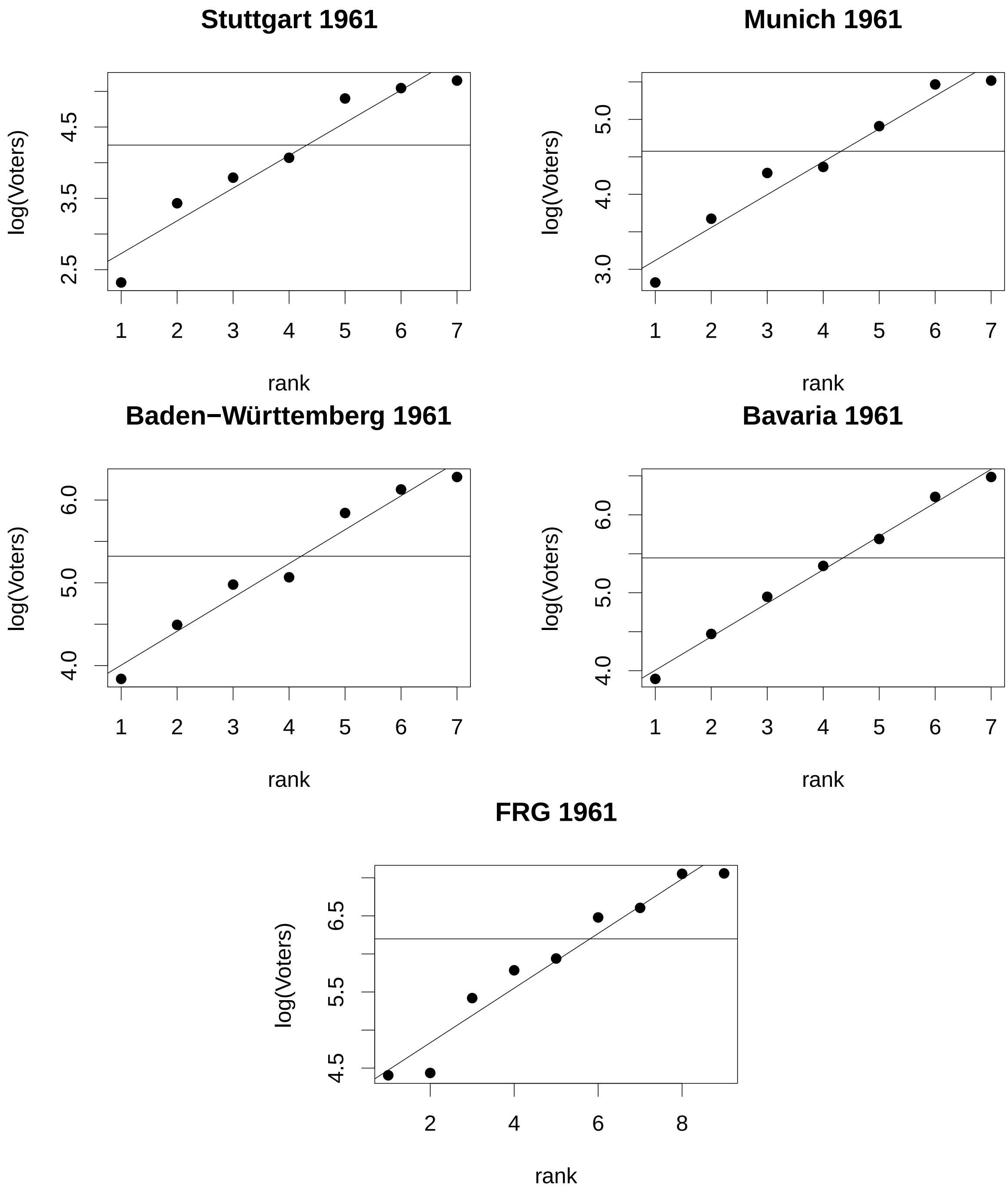


Figure 14: Election FRG, 1961 (bullets: data, line: linear fit).

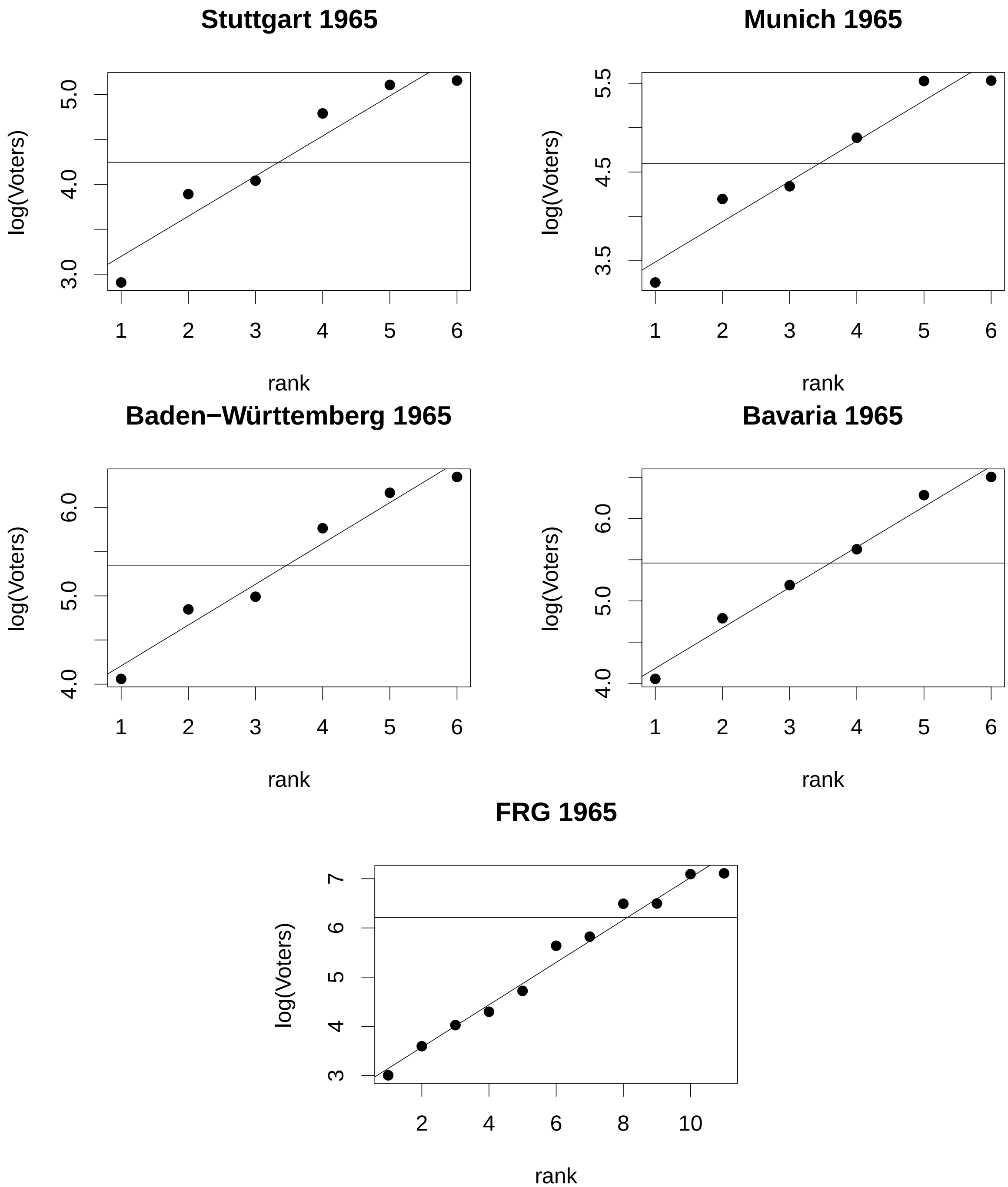


Figure 15: Election FRG, 1965 (bullets: data, line: linear fit).

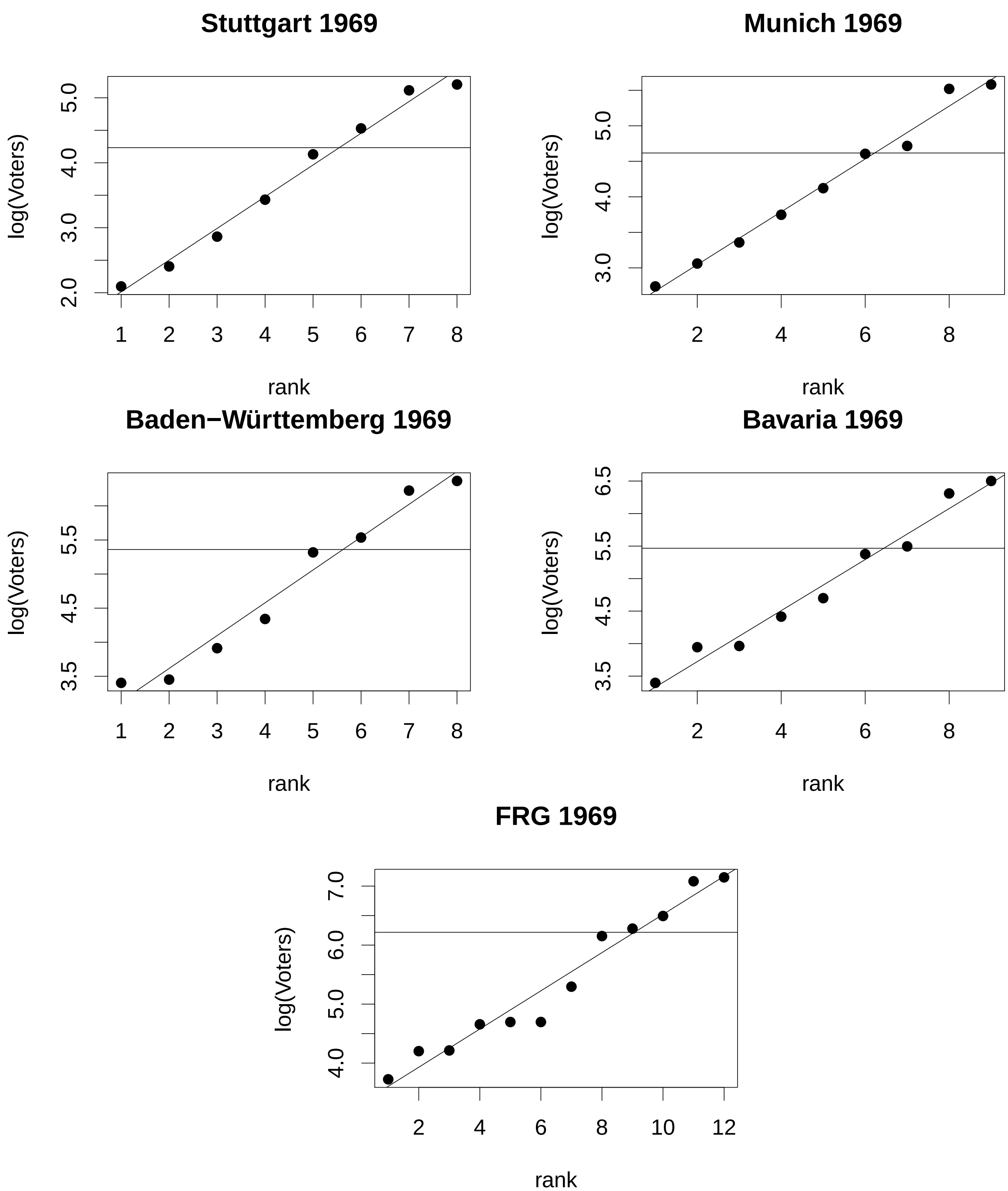


Figure 16: Election FRG, 1969 (bullets: data, line: linear fit).

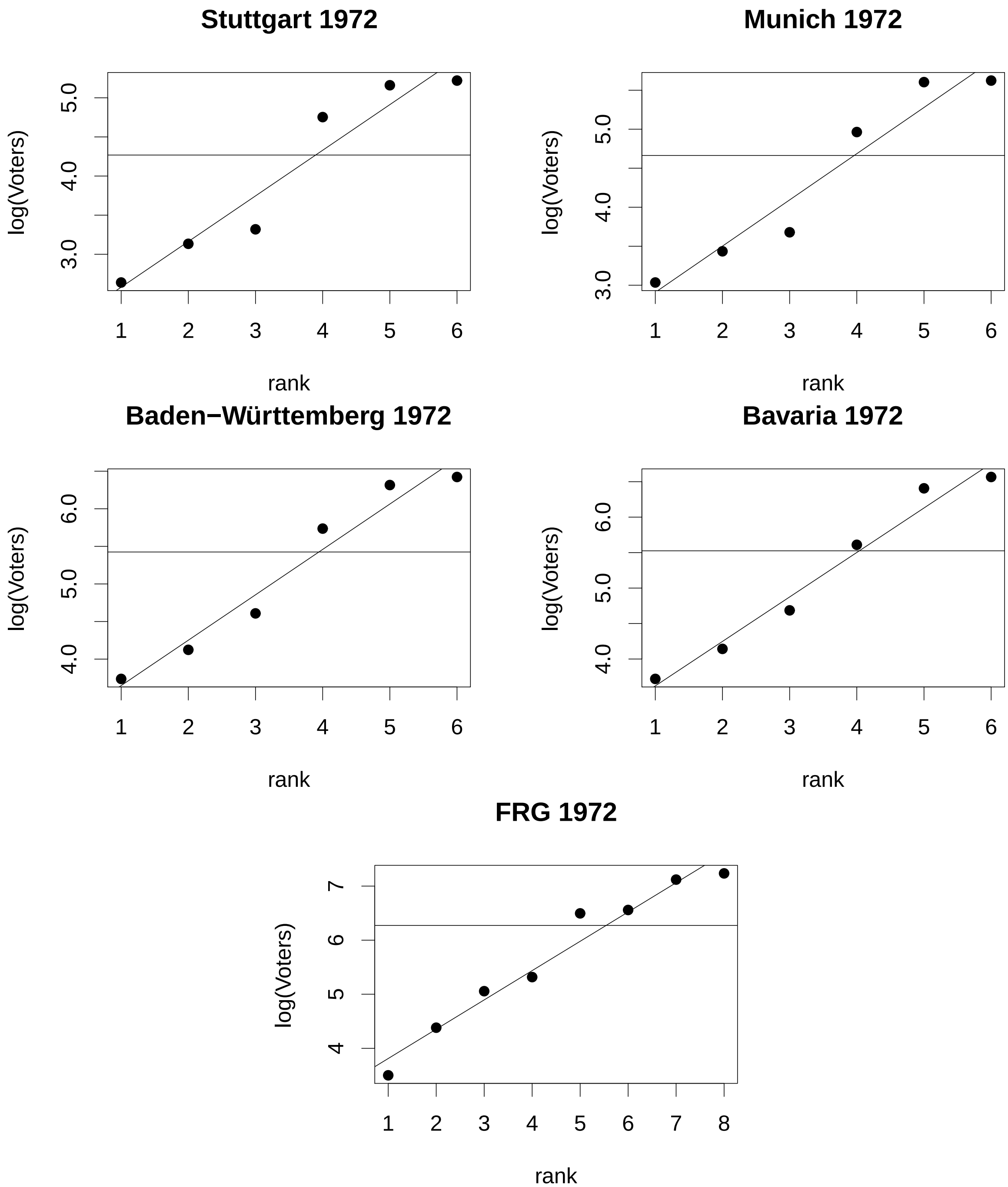


Figure 17: Election FRG, 1972 (bullets: data, line: linear fit).

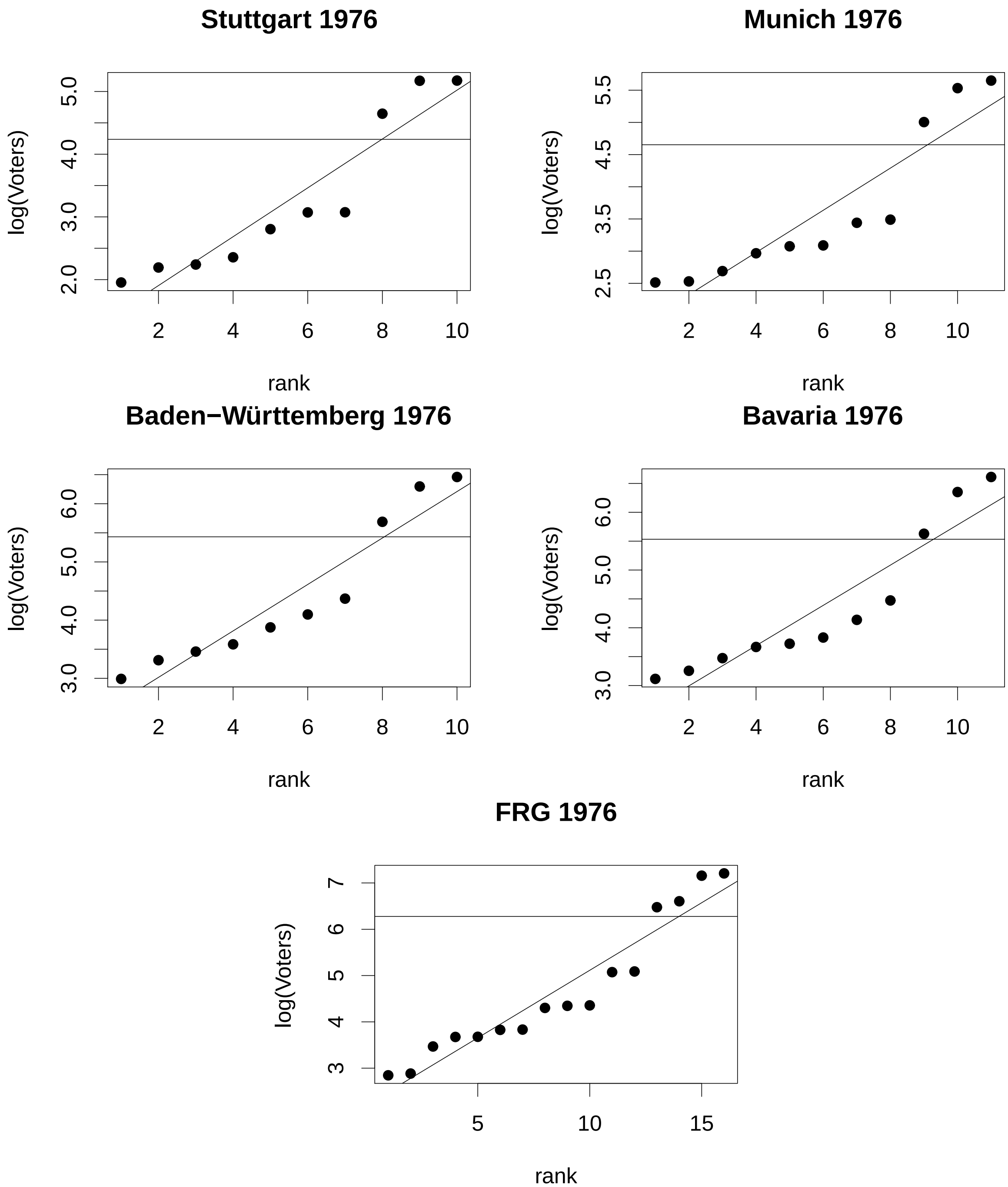


Figure 18: Election FRG, 1976 (bullets: data, line: linear fit).

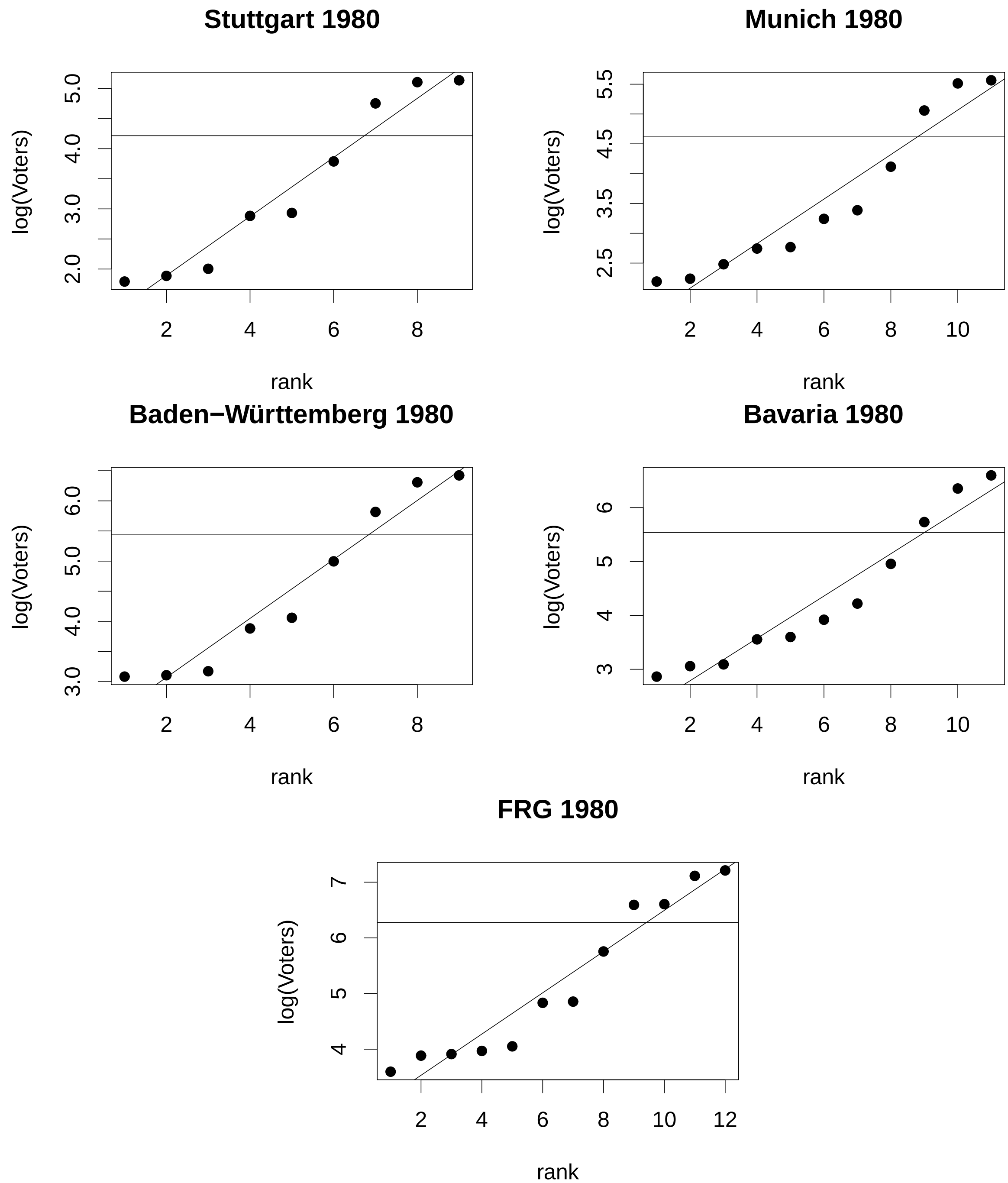


Figure 19: Election FRG, 1980 (bullets: data, line: linear fit).

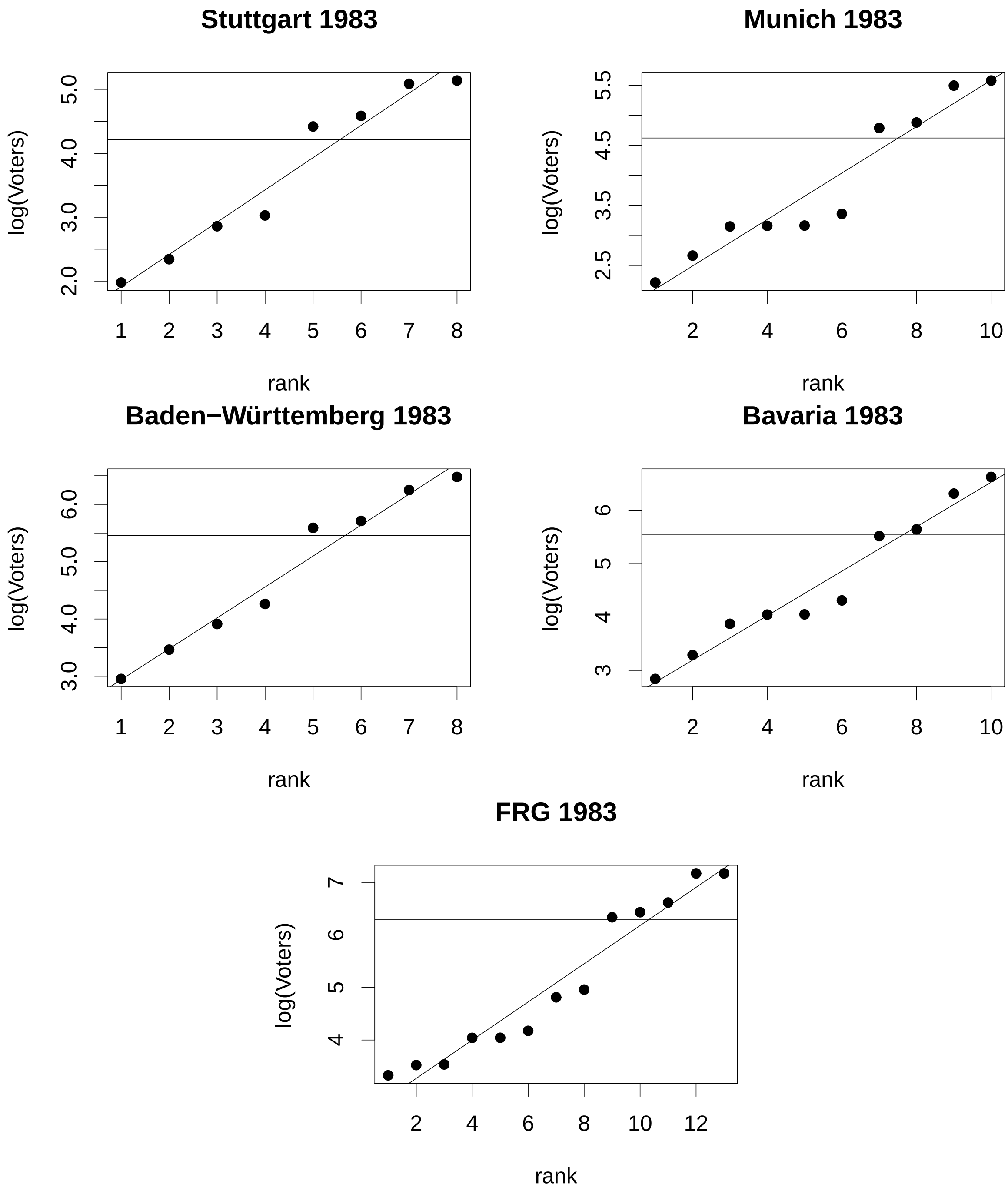


Figure 20: Election FRG, 1983 (bullets: data, line: linear fit).

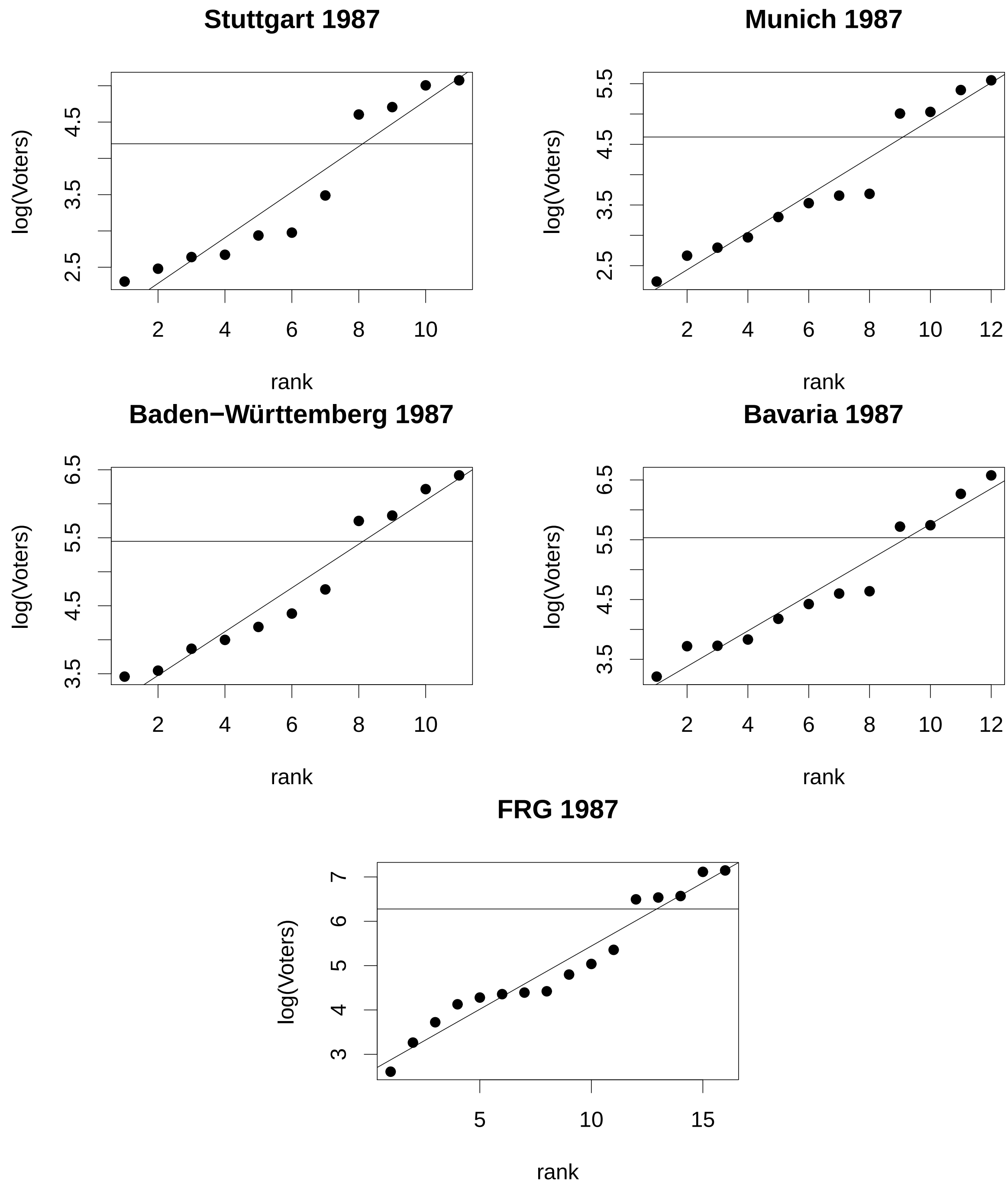


Figure 21: Election FRG, 1987 (bullets: data, line: linear fit).

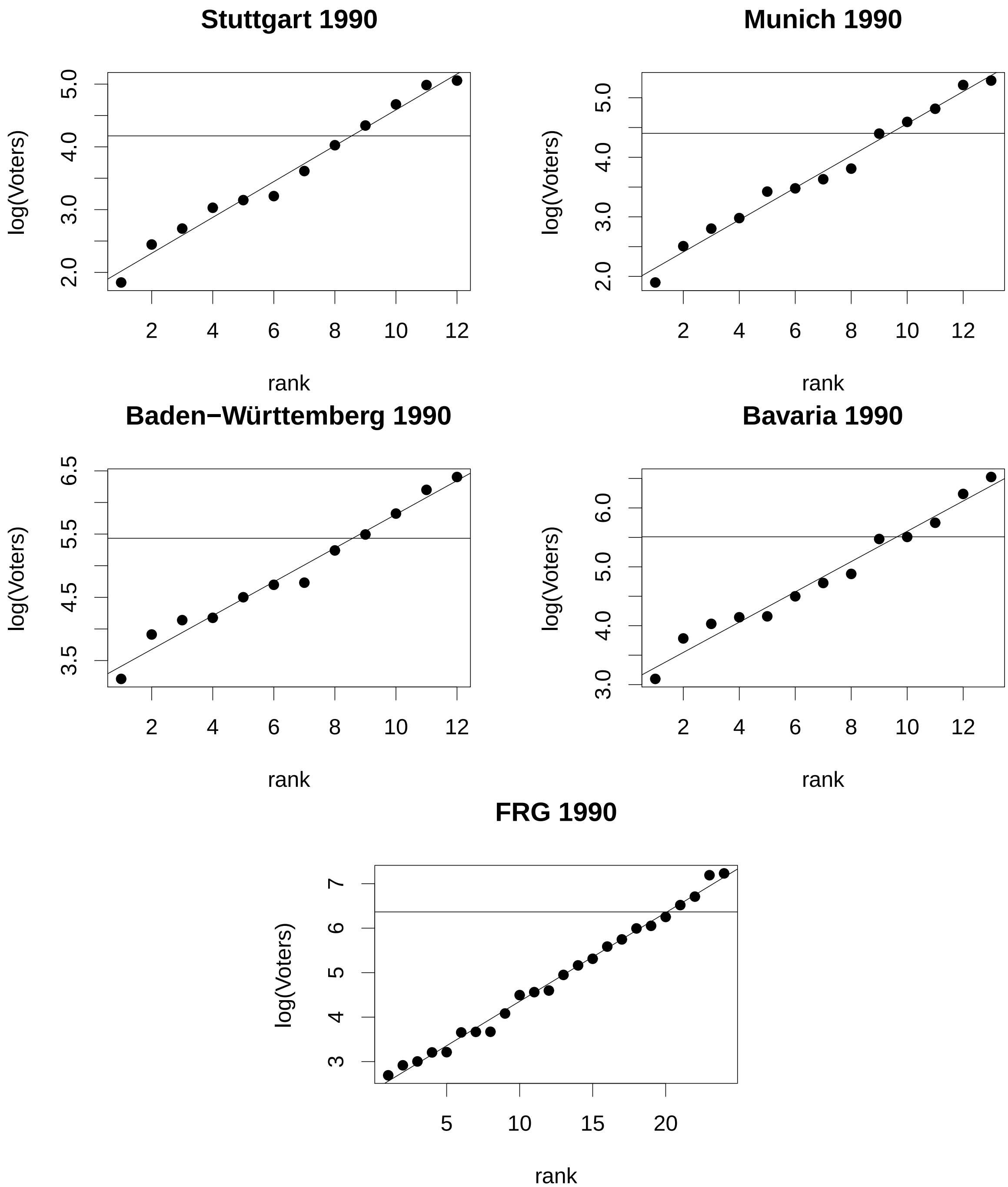


Figure 22: Election FRG, 1990 (bullets: data, line: linear fit).

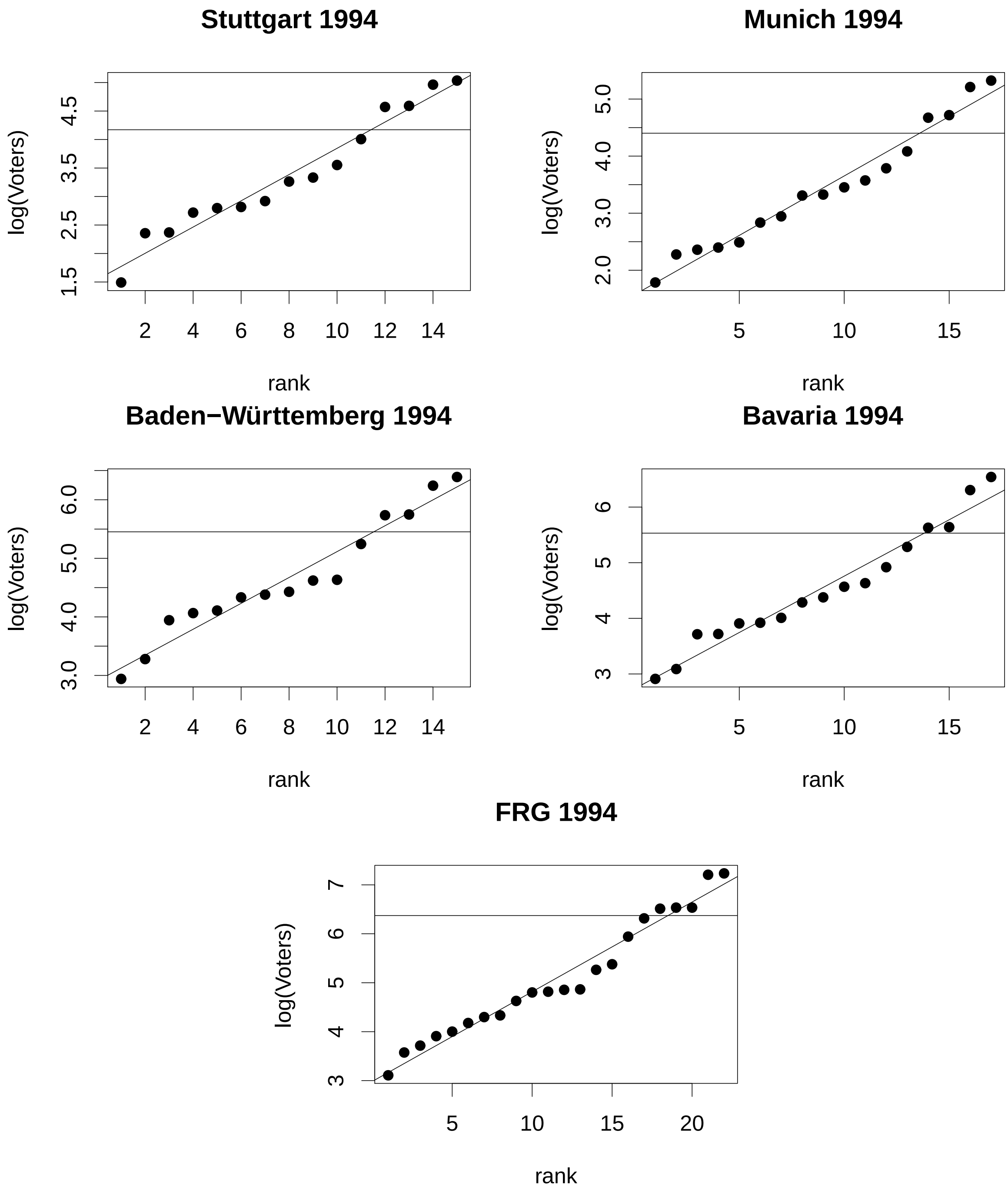


Figure 23: Election FRG, 1994 (bullets: data, line: linear fit).

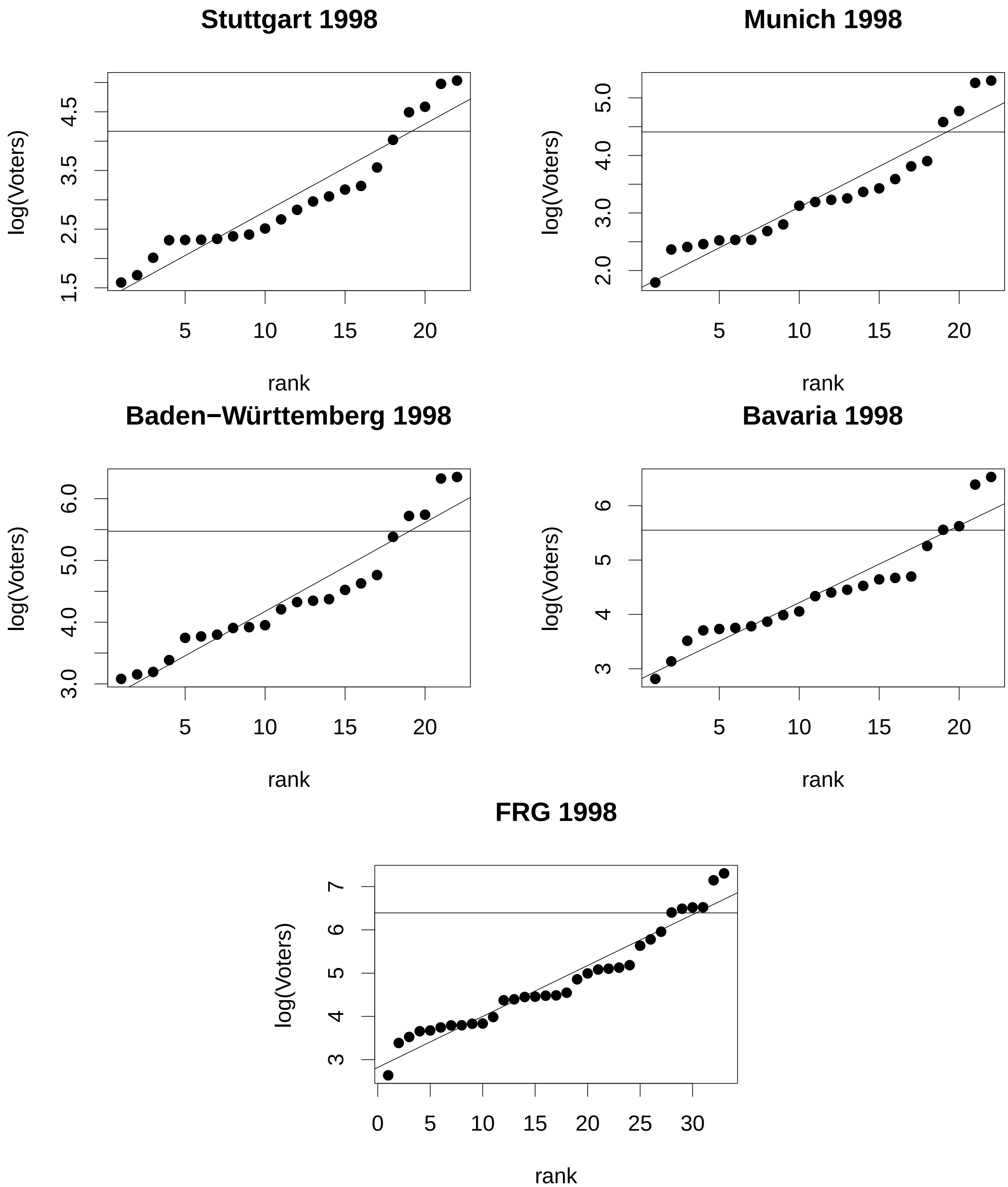


Figure 24: Election FRG, 1998 (bullets: data, line: linear fit).

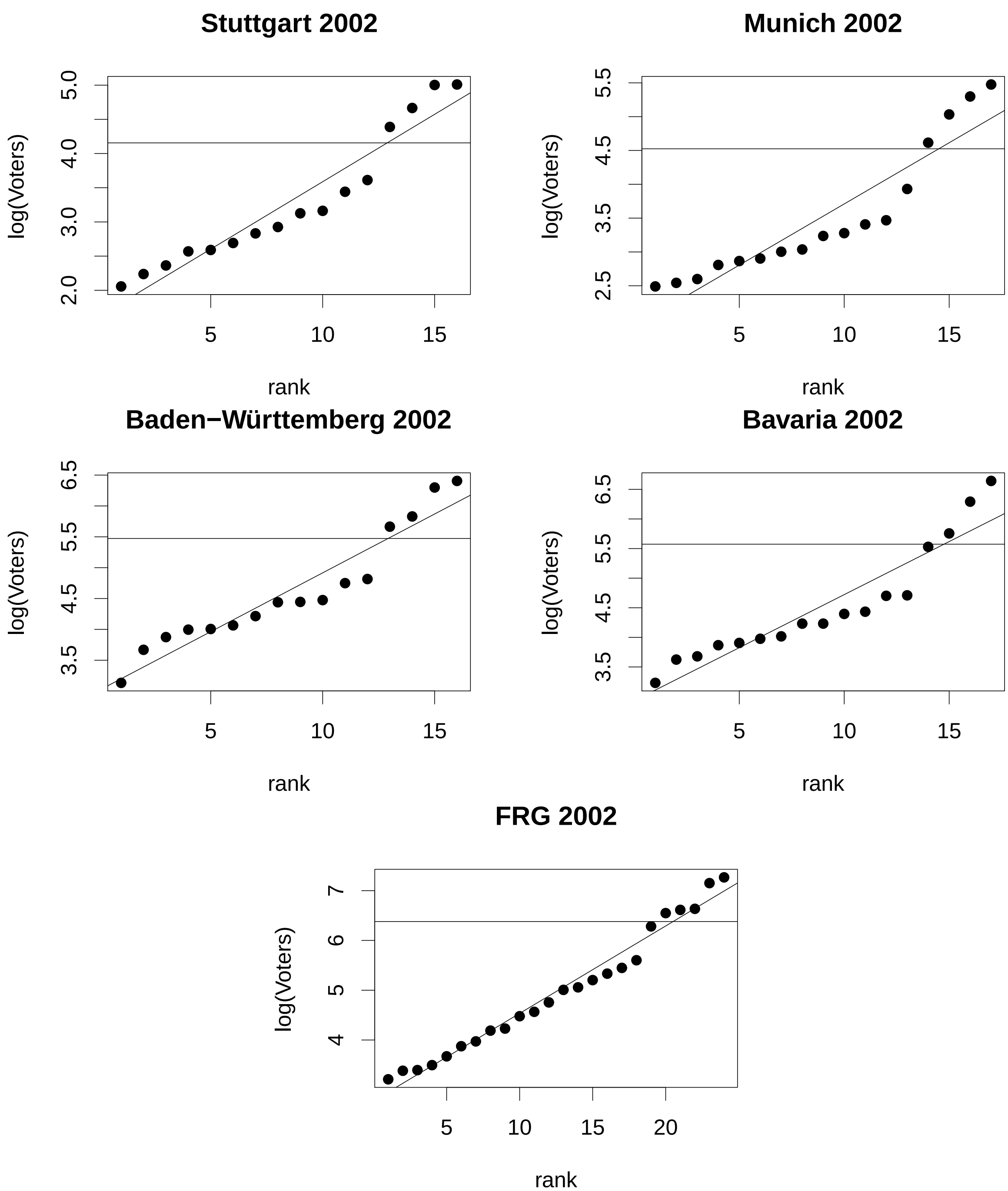


Figure 25: Election FRG, 2002 (bullets: data, line: linear fit).

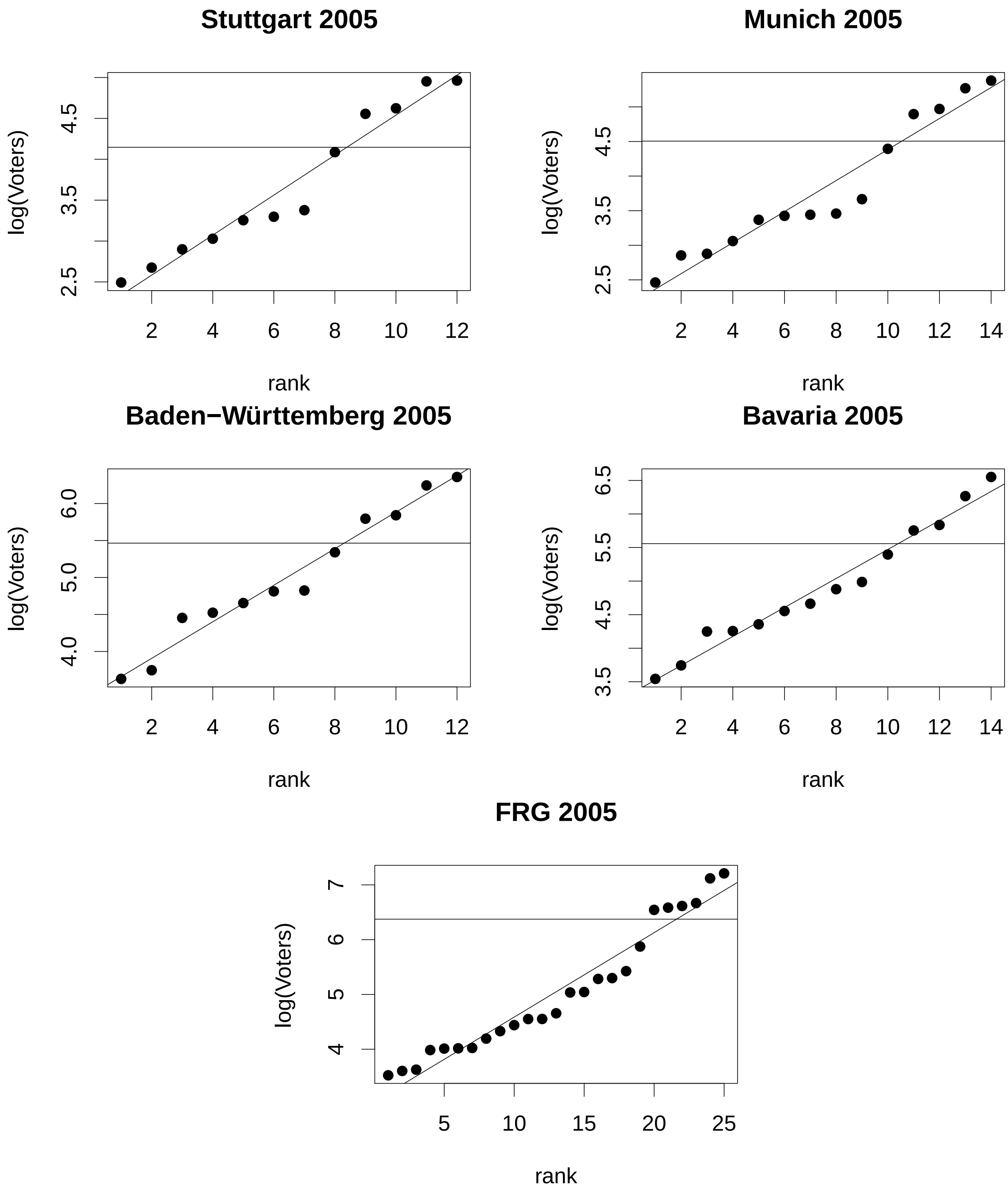


Figure 26: Election FRG, 2005 (bullets: data, line: linear fit).

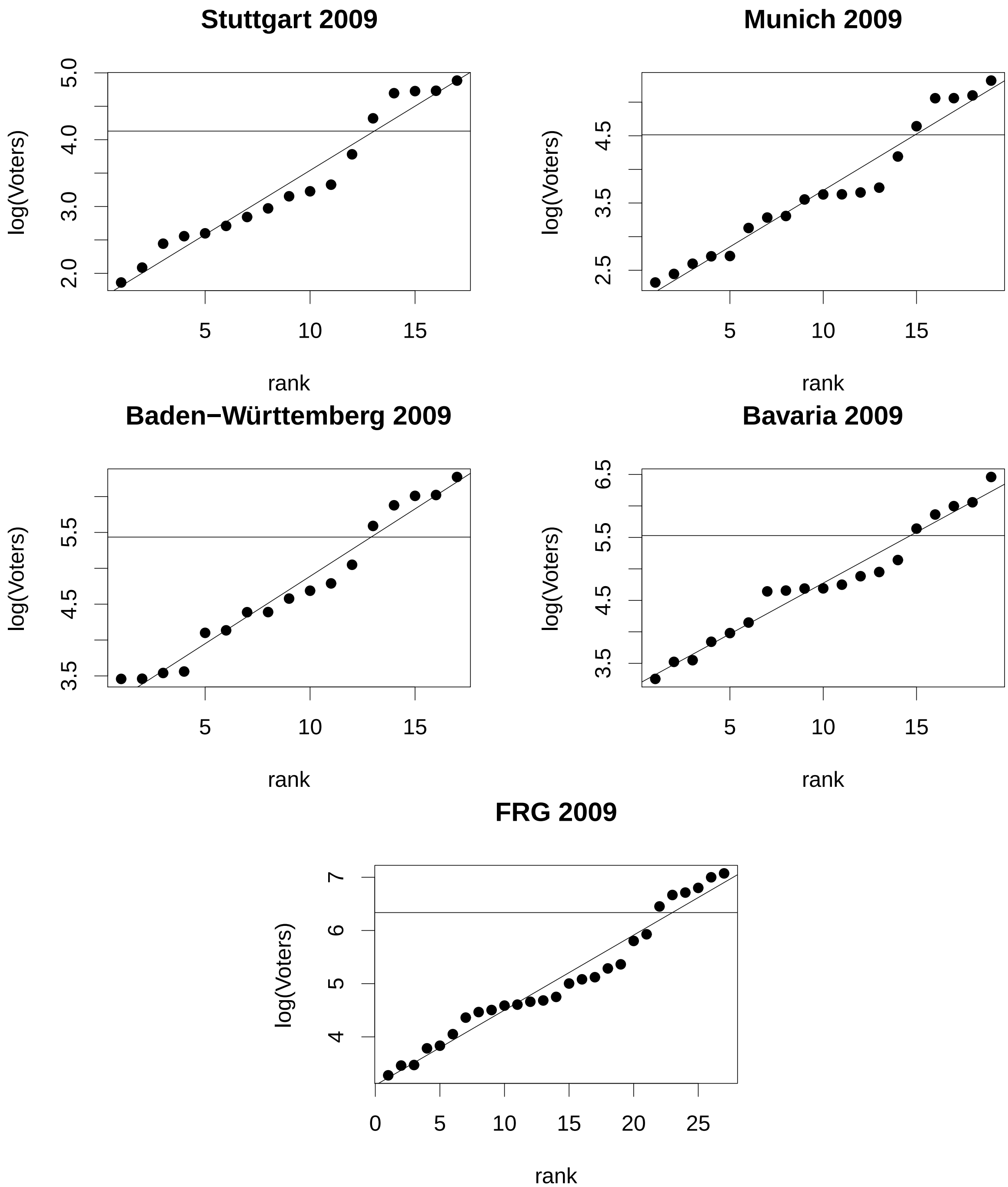


Figure 27: Election FRG, 2009 (bullets: data, line: linear fit).

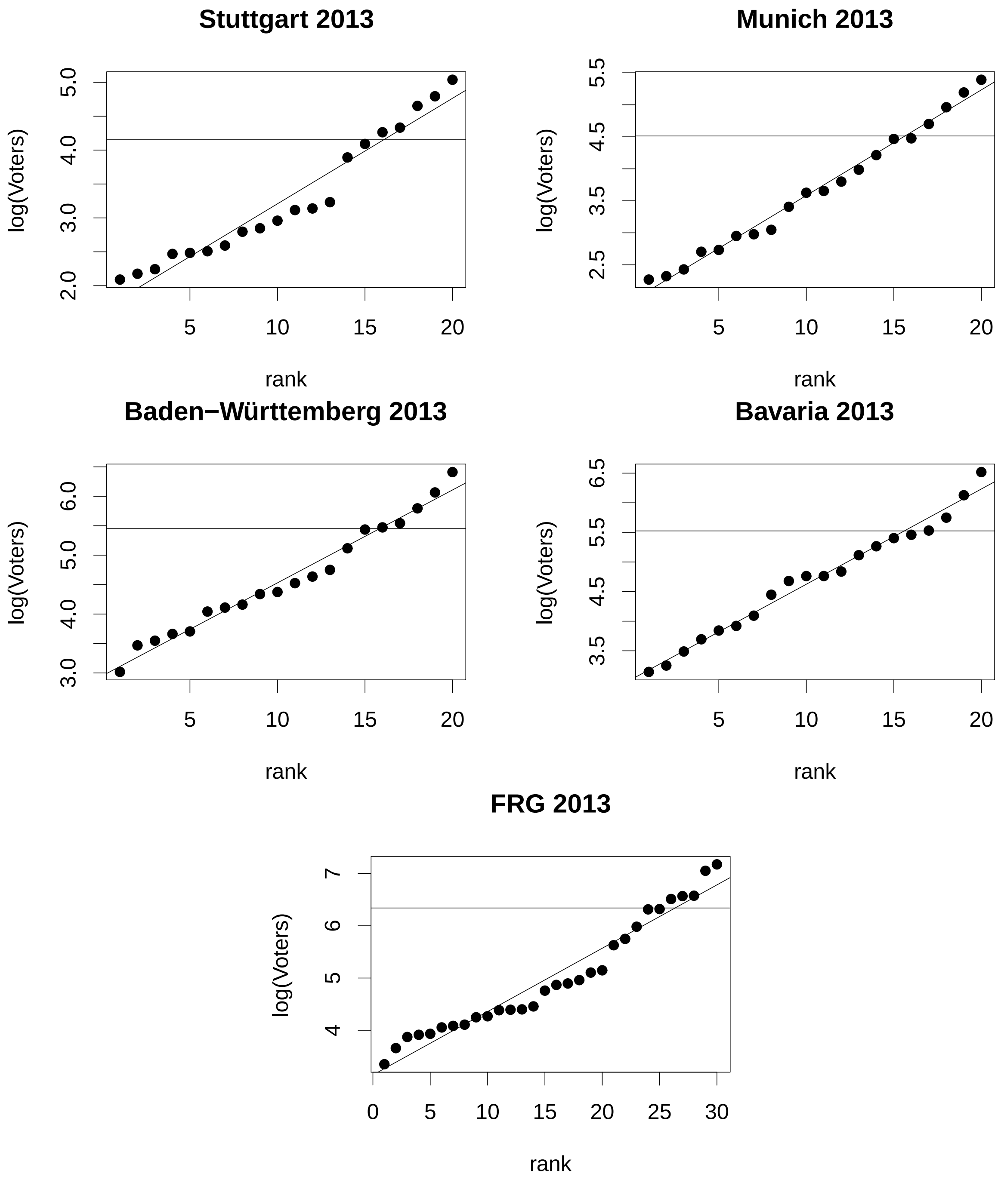


Figure 28: Election FRG, 2013 (bullets: data, line: linear fit).

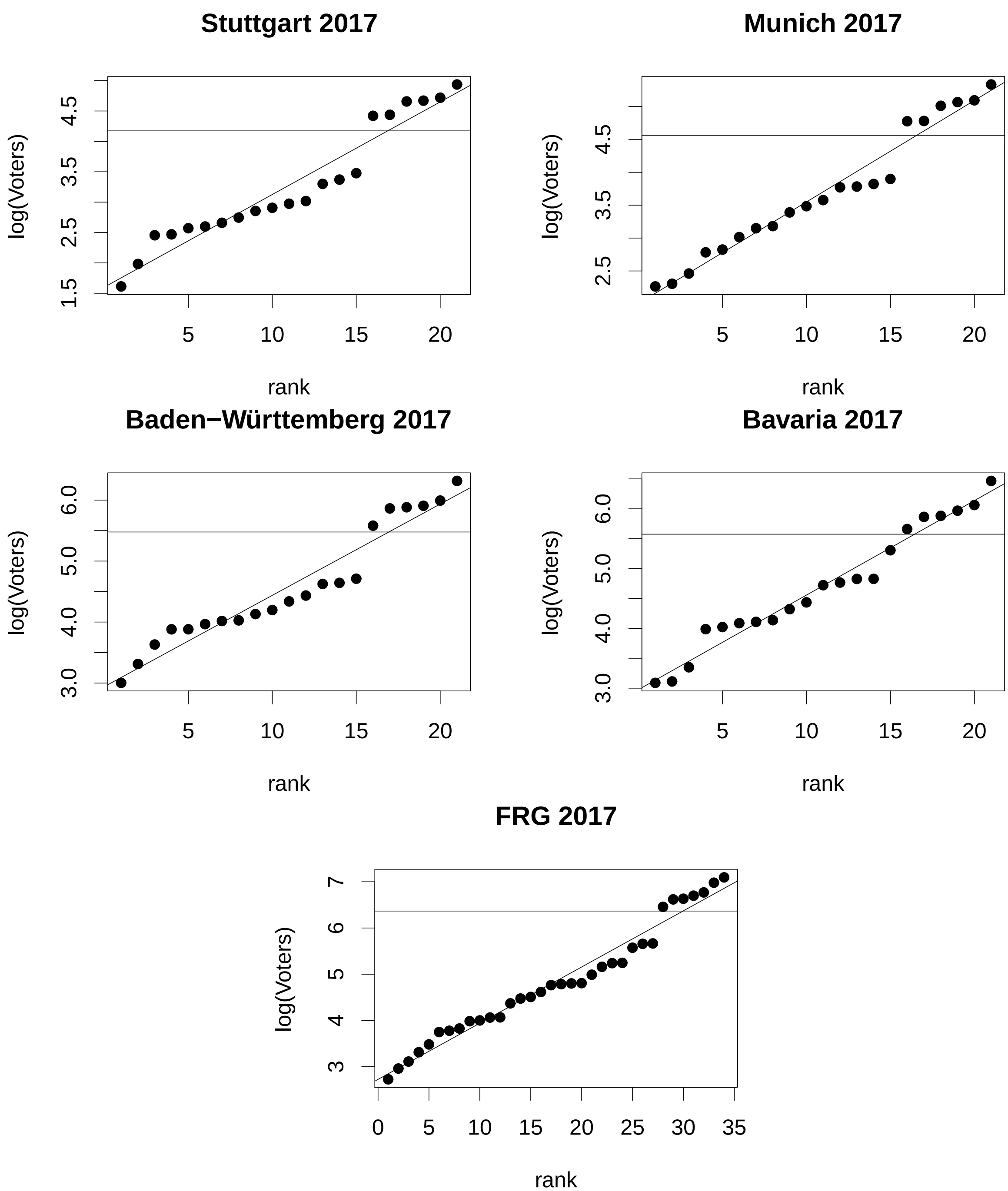


Figure 29: Election FRG, 2017 (bullets: data, line: linear fit).

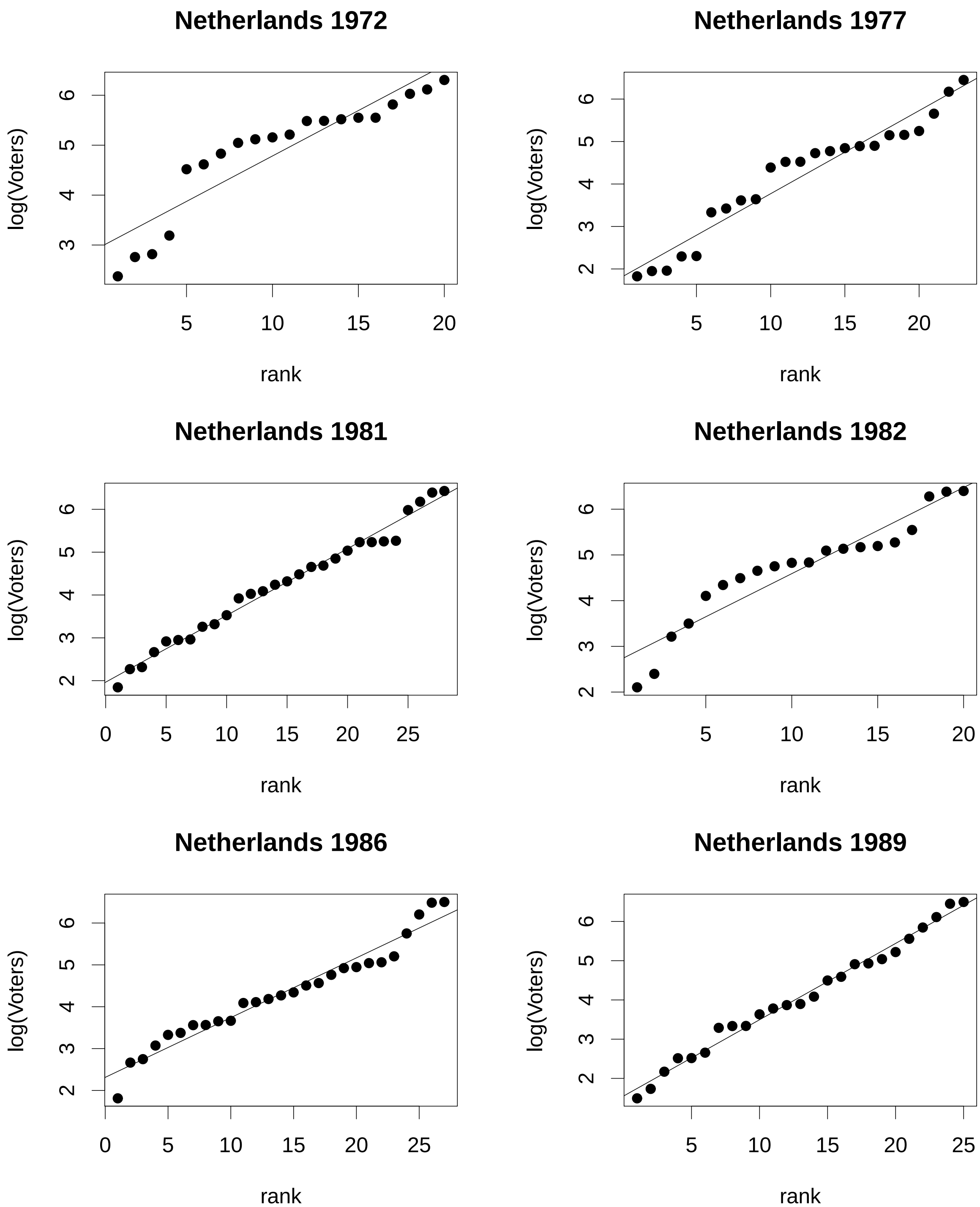


Figure 30: Election in the Netherlands (bullets: data, line: linear fit)

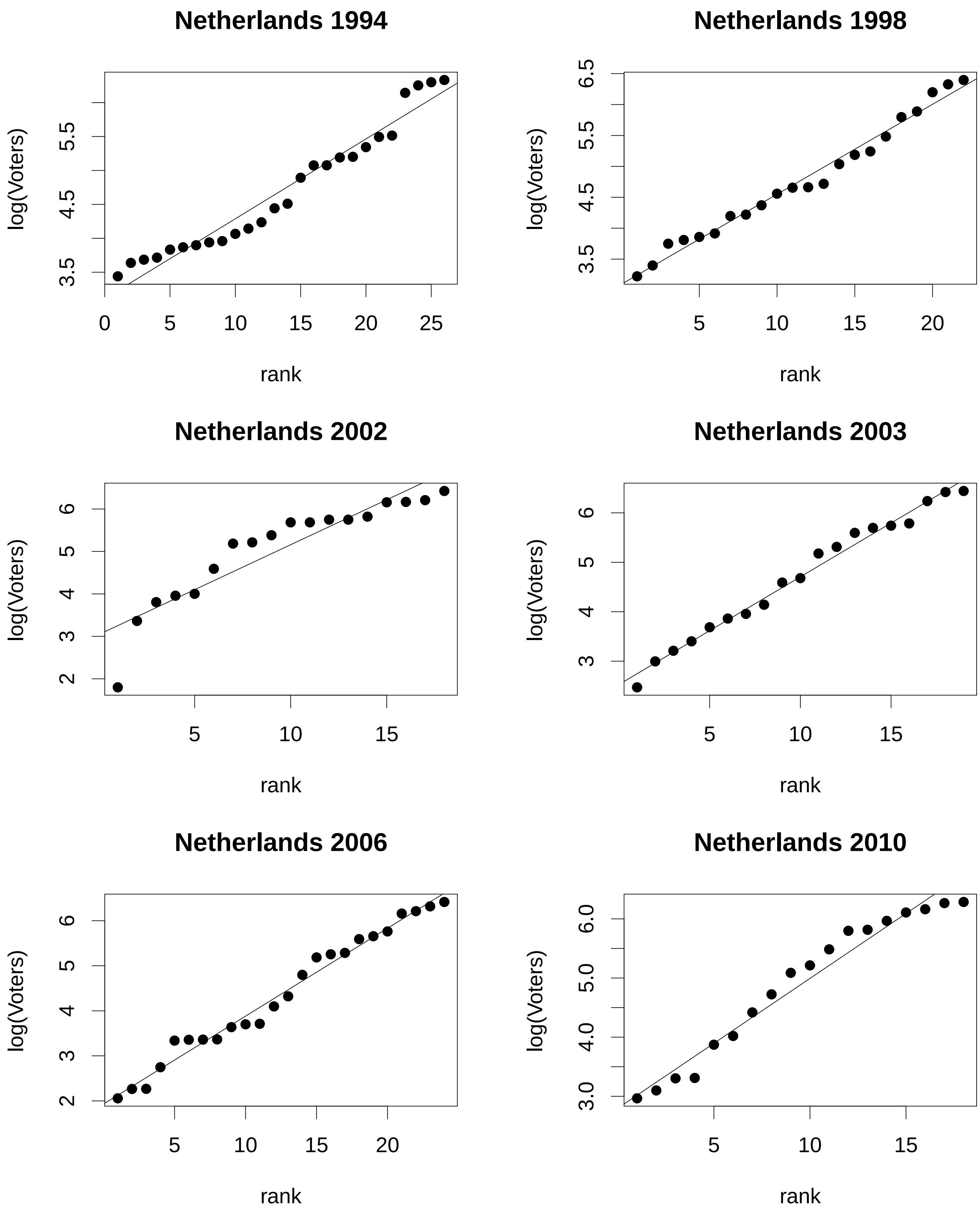


Figure 31: Election in the Netherlands (bullets: data, line: linear fit).

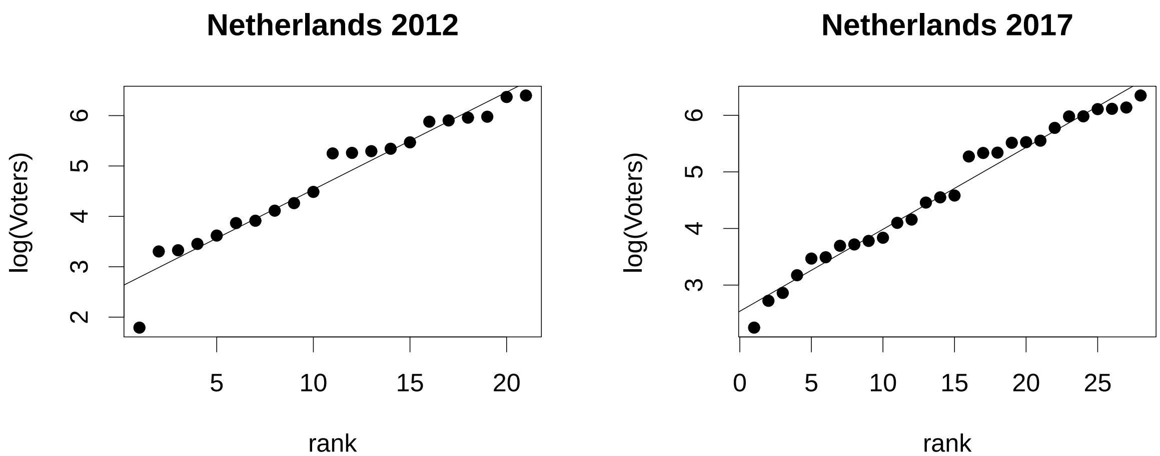


Figure 32: Election in the Netherlands (bullets: data, line: linear fit).

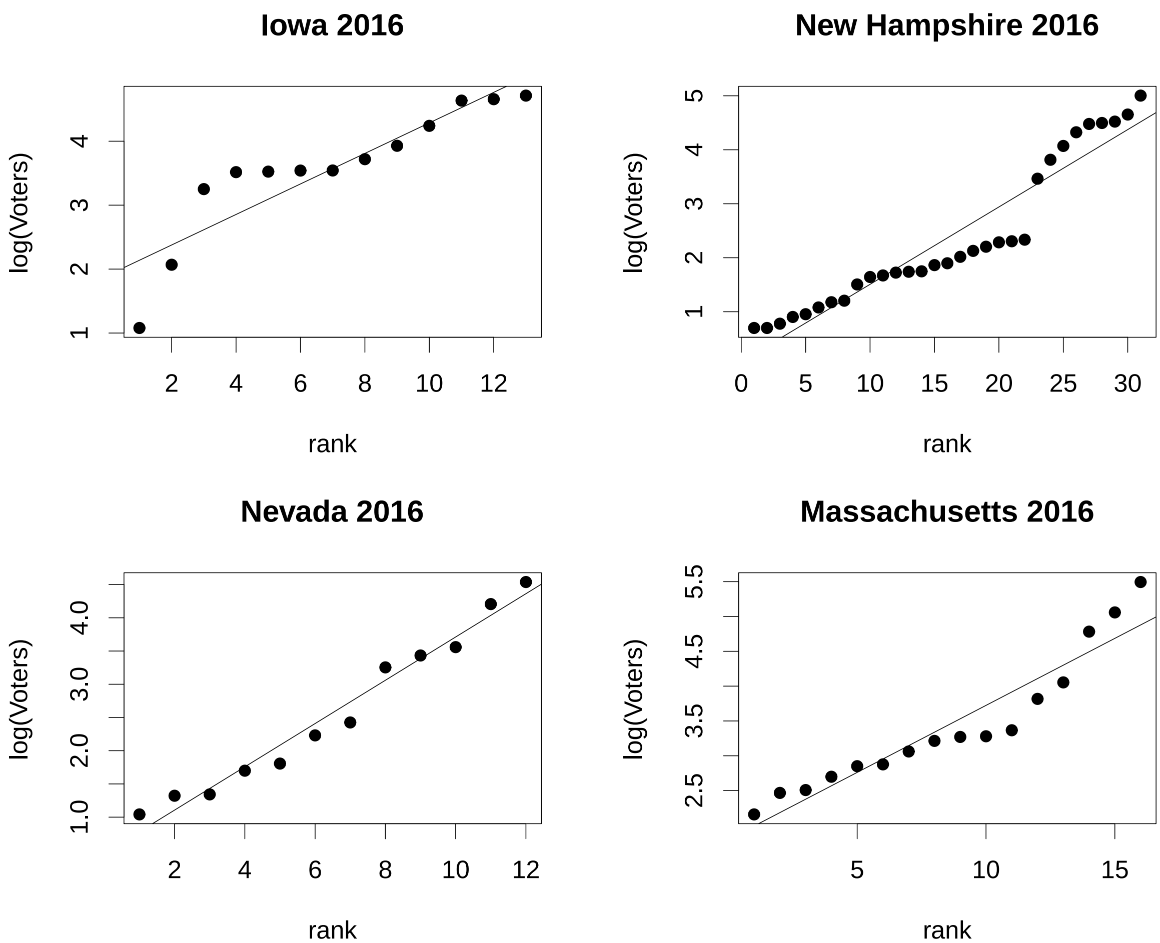


Figure 33: Election US (republican primaries), 2016 (bullets: data, line: linear fit).

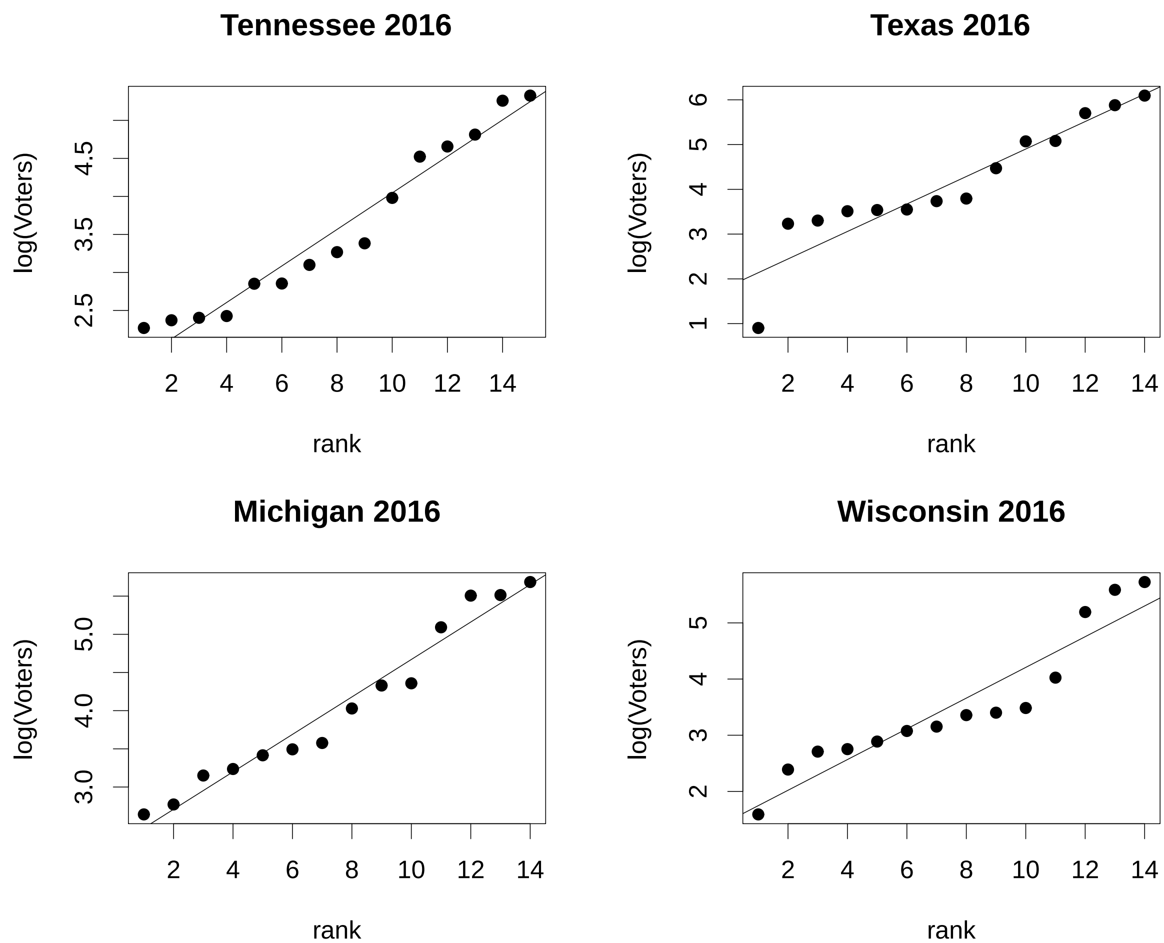


Figure 34: Election US (republican primaries), 2016 (bullets: data, line: linear fit).

6.1 Parameter values of the linear fit

Note that – for a given election – the slope is rather independent on the number of voters, provided that the number of parties (candidates) are the same for organizational units of different size. If we go from cities to states to the country (more and more voters are involved), mainly the intercept changes.

country	unit	year	#parties	#voters	intercept	slope	R^2
FRG	Stuttgart	1949	6	224203	3.552	0.245	0.68
FRG	Munich	1949	6	503822	4.5	0.11	0.951
FRG	Bavaria	1949	7	4824052	4.112	0.347	0.606
FRG	Baden-Württemberg	1949	7	2745453	3.556	0.4	0.813
FRG	FRG	1949	15	23732398	4.497	0.16	0.895
FRG	Stuttgart	1953	8	303222	3.167	0.24	0.922
FRG	Munich	1953	10	589476	3.12	0.223	0.979
FRG	Bavaria	1953	10	5178145	4.077	0.218	0.956
FRG	Baden-Württemberg	1953	8	3588131	4.15	0.248	0.937
FRG	FRG	1953	13	27551272	4.582	0.182	0.973
FRG	Stuttgart	1957	8	344679	1.987	0.429	0.958
FRG	Munich	1957	11	690061	2.148	0.301	0.965
FRG	Bavaria	1957	11	5380289	3.126	0.288	0.965
FRG	Baden-Württemberg	1957	8	3907840	3.282	0.387	0.987
FRG	FRG	1957	13	29905428	3.572	0.278	0.984
FRG	Stuttgart	1961	7	353395	2.268	0.458	0.911
FRG	Munich	1961	7	751046	2.679	0.439	0.943
FRG	Bavaria	1961	7	5596615	3.576	0.43	0.991
FRG	Baden-Württemberg	1961	7	4189163	3.596	0.409	0.959
FRG	FRG	1961	9	31550901	4.116	0.359	0.943
FRG	Stuttgart	1965	6	351760	2.752	0.447	0.9
FRG	Munich	1965	6	791871	3.03	0.455	0.927
FRG	Bavaria	1965	6	5772593	3.691	0.491	0.98
FRG	Baden-Württemberg	1965	6	4452227	3.745	0.462	0.953
FRG	FRG	1965	11	32620442	2.715	0.431	0.977
FRG	Stuttgart	1969	8	341860	1.525	0.488	0.983
FRG	Munich	1969	9	827541	2.301	0.372	0.984
FRG	Bavaria	1969	9	5860298	2.939	0.392	0.973
FRG	Baden-Württemberg	1969	8	4584766	2.652	0.481	0.963
FRG	FRG	1969	12	32966024	3.282	0.324	0.956
FRG	Stuttgart	1972	6	371579	1.996	0.583	0.902
FRG	Munich	1972	6	922487	2.317	0.592	0.917
FRG	Bavaria	1972	6	6708665	2.993	0.627	0.968
FRG	Baden-Württemberg	1972	6	5322133	3.044	0.604	0.946
FRG	FRG	1972	8	37459750	3.269	0.542	0.951

country	unit	year	#parties	#voters	intercept	slope	R^2
FRG	Stuttgart	1976	10	345308	1.126	0.39	0.866
FRG	Munich	1976	11	897795	1.669	0.328	0.825
FRG	Bavaria	1976	11	6822919	2.296	0.349	0.853
FRG	Baden-Württemberg	1976	10	5405534	2.219	0.399	0.888
FRG	FRG	1976	16	37822500	2.196	0.292	0.908
FRG	Stuttgart	1980	9	328751	0.912	0.491	0.943
FRG	Munich	1980	11	827190	1.332	0.373	0.915
FRG	Bavaria	1980	11	6896186	2.002	0.393	0.924
FRG	Baden-Württemberg	1980	9	5454040	2.092	0.489	0.936
FRG	FRG	1980	12	37938981	2.791	0.37	0.929
FRG	Stuttgart	1983	8	329103	1.406	0.506	0.943
FRG	Munich	1983	10	840650	1.716	0.387	0.912
FRG	Bavaria	1983	10	7077256	2.359	0.417	0.951
FRG	Baden-Württemberg	1983	8	5722585	2.4	0.539	0.962
FRG	FRG	1983	13	38940687	2.543	0.364	0.94
FRG	Stuttgart	1987	11	317464	1.65	0.314	0.904
FRG	Munich	1987	12	835534	1.814	0.308	0.94
FRG	Bavaria	1987	12	6846746	2.786	0.297	0.943
FRG	Baden-Württemberg	1987	11	5608973	2.832	0.322	0.945
FRG	FRG	1987	16	37867319	2.589	0.285	0.948
FRG	Stuttgart	1990	12	299515	1.733	0.286	0.983
FRG	Munich	1990	13	505886	1.87	0.27	0.982
FRG	Bavaria	1990	13	6468521	3.031	0.257	0.972
FRG	Baden-Württemberg	1990	12	5439352	3.142	0.267	0.975
FRG	FRG	1990	24	46455772	2.36	0.199	0.993
FRG	Stuttgart	1994	15	297117	1.543	0.23	0.954
FRG	Munich	1994	17	504335	1.569	0.208	0.967
FRG	Bavaria	1994	17	6799366	2.732	0.203	0.955
FRG	Baden-Württemberg	1994	15	5668824	2.905	0.221	0.942
FRG	FRG	1994	22	47105174	2.983	0.183	0.964
FRG	Stuttgart	1998	22	294613	1.303	0.15	0.905
FRG	Munich	1998	22	512338	1.687	0.141	0.906
FRG	Bavaria	1998	22	7085122	2.799	0.142	0.909
FRG	Baden-Württemberg	1998	22	5945364	2.736	0.144	0.917
FRG	FRG	1998	33	49308512	2.821	0.118	0.942

country	unit	year	#parties	#voters	intercept	slope	R^2
FRG	Stuttgart	2002	16	286544	1.617	0.197	0.914
FRG	Munich	2002	17	669491	1.9	0.181	0.854
FRG	Bavaria	2002	17	7495435	2.929	0.179	0.868
FRG	Baden-Württemberg	2002	16	5939859	3.009	0.191	0.902
FRG	FRG	2002	24	47996480	2.779	0.176	0.975
FRG	Stuttgart	2005	12	280490	2.094	0.245	0.95
FRG	Munich	2005	14	640630	2.141	0.224	0.929
FRG	Bavaria	2005	14	7222308	3.31	0.216	0.97
FRG	Baden-Württemberg	2005	12	5822447	3.413	0.247	0.966
FRG	FRG	2005	25	47287988	3.044	0.154	0.946
FRG	Stuttgart	2009	17	269104	1.619	0.192	0.952
FRG	Munich	2009	19	652946	2.01	0.168	0.955
FRG	Bavaria	2009	19	6768352	3.158	0.162	0.964
FRG	Baden-Württemberg	2009	17	5442089	3.01	0.188	0.967
FRG	FRG	2009	27	43371190	3.09	0.141	0.96
FRG	Stuttgart	2013	20	284541	1.651	0.156	0.944
FRG	Munich	2013	20	650216	1.933	0.165	0.989
FRG	Bavaria	2013	20	6695559	3.015	0.161	0.982
FRG	Baden-Württemberg	2013	20	5642019	2.953	0.158	0.976
FRG	FRG	2013	30	43726856	3.146	0.121	0.947
FRG	Stuttgart	2017	21	298012	1.601	0.152	0.925
FRG	Munich	2017	21	722141	2.007	0.154	0.966
FRG	Bavaria	2017	21	7519739	2.977	0.158	0.968
FRG	Baden-Württemberg	2017	21	5992968	2.941	0.15	0.928
FRG	FRG	2017	34	46515492	2.725	0.122	0.969
F	France	2007	12	36719396	5.02	0.167	0.924
F	France	2012	10	35883209	4.844	0.238	0.965
F	France	2017	11	36054394	4.792	0.217	0.935
F	Paris	2017	11	1076559	2.883	0.254	0.977
F	Ile de France	2017	11	5631456	3.833	0.234	0.966
F	Lyon	2017	11	234507	2.383	0.24	0.964
F	Rhone-Alpes	2017	11	4187716	3.853	0.218	0.936
F	Marseille	2017	11	366083	2.595	0.238	0.949
F	Cote d'Azur	2017	11	2750937	3.525	0.233	0.955
NL	Netherlands	2017	28	10516041	2.537	0.145	0.975
NL	Netherlands	2012	21	9424235	2.599	0.193	0.938
NL	Netherlands	2010	18	9416001	2.799	0.219	0.96
NL	Netherlands	2006	24	9838683	1.931	0.195	0.981
NL	Netherlands	2003	19	9654475	2.528	0.218	0.985
NL	Netherlands	2002	18	10545916	3.041	0.212	0.836
NL	Netherlands	1998	22	8607787	3.095	0.145	0.982
NL	Netherlands	1994	26	8981556	3.112	0.118	0.953
NL	Netherlands	1989	25	8893302	1.548	0.195	0.987
NL	Netherlands	1986	27	9172159	2.313	0.143	0.951
NL	Netherlands	1982	20	8236516	2.705	0.188	0.892
NL	Netherlands	1981	28	8690837	1.966	0.156	0.984
NL	Netherlands	1977	23	5665334	1.816	0.196	0.938
NL	Netherlands	1972	20	7394045	2.964	0.182	0.833

country	unit	year	#parties	#voters	intercept	slope	R^2
US	Iowa	2016	13	186872	1.9	0.239	0.793
US	New Hampshire	2016	31	284149	0.074	0.143	0.904
US	Nevada	2016	12	59336	0.455	0.326	0.968
US	Massachusetts	2016	16	516932	1.799	0.192	0.886
US	Tennessee	2016	15	553752	1.644	0.24	0.951
US	Texas	2016	14	2796618	1.831	0.307	0.863
US	Michigan	2016	14	1323589	2.218	0.245	0.959
US	Wisconsin	2016	14	1099469	1.474	0.273	0.875

7 Elections in semilogarithmic representation with model simulations

In this section, we show a boxplot of 100 realization of the model (n, K and z adapted) together with the data from the corresponding election according to the algorithm described in section 2.4.4. The number of voters n and the number of candidates/parties K are directly taken from the data, the relative minimal group size z is estimated according to the estimator described at page 17 (this supplement).

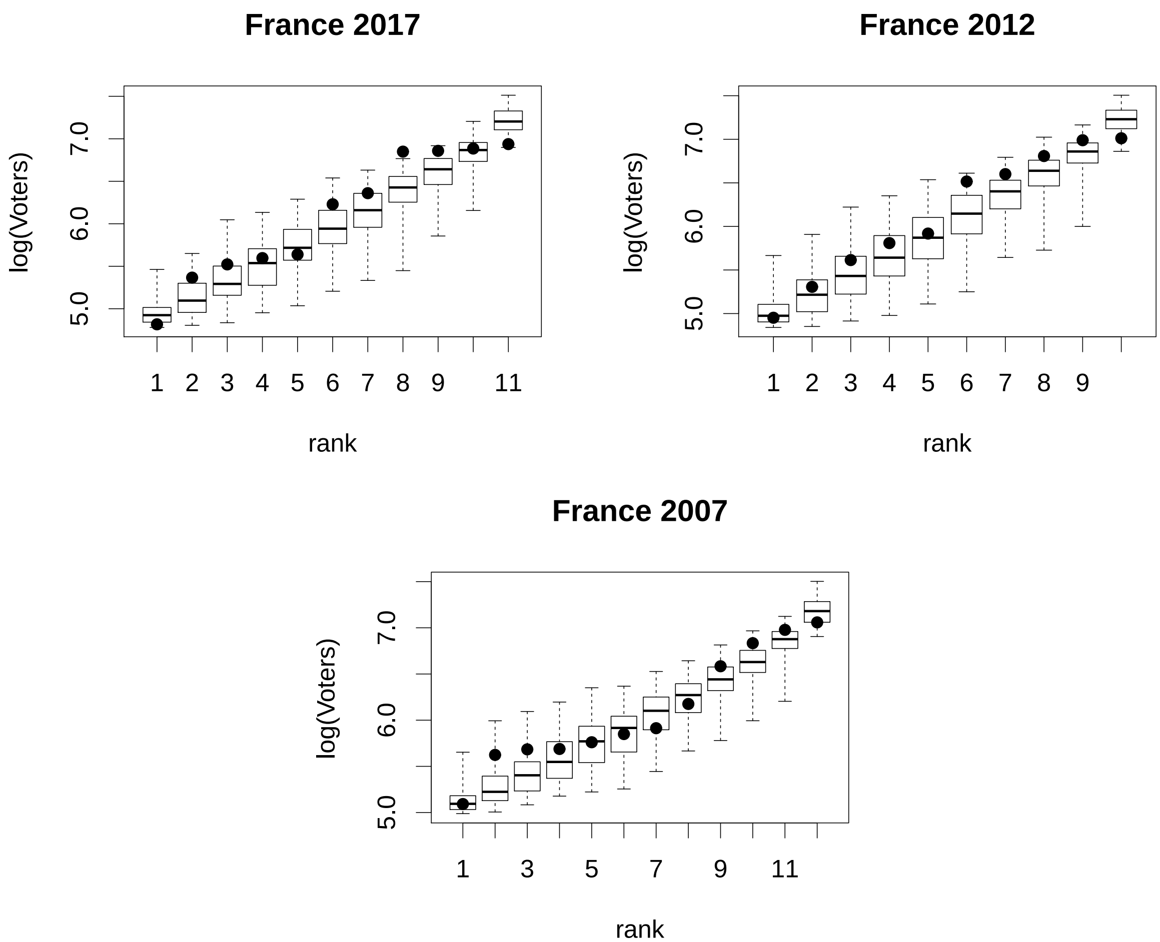


Figure 35: Election France, 2017, 1012, 2007 (boxplot of 100 realizations og the model, bullets: data).

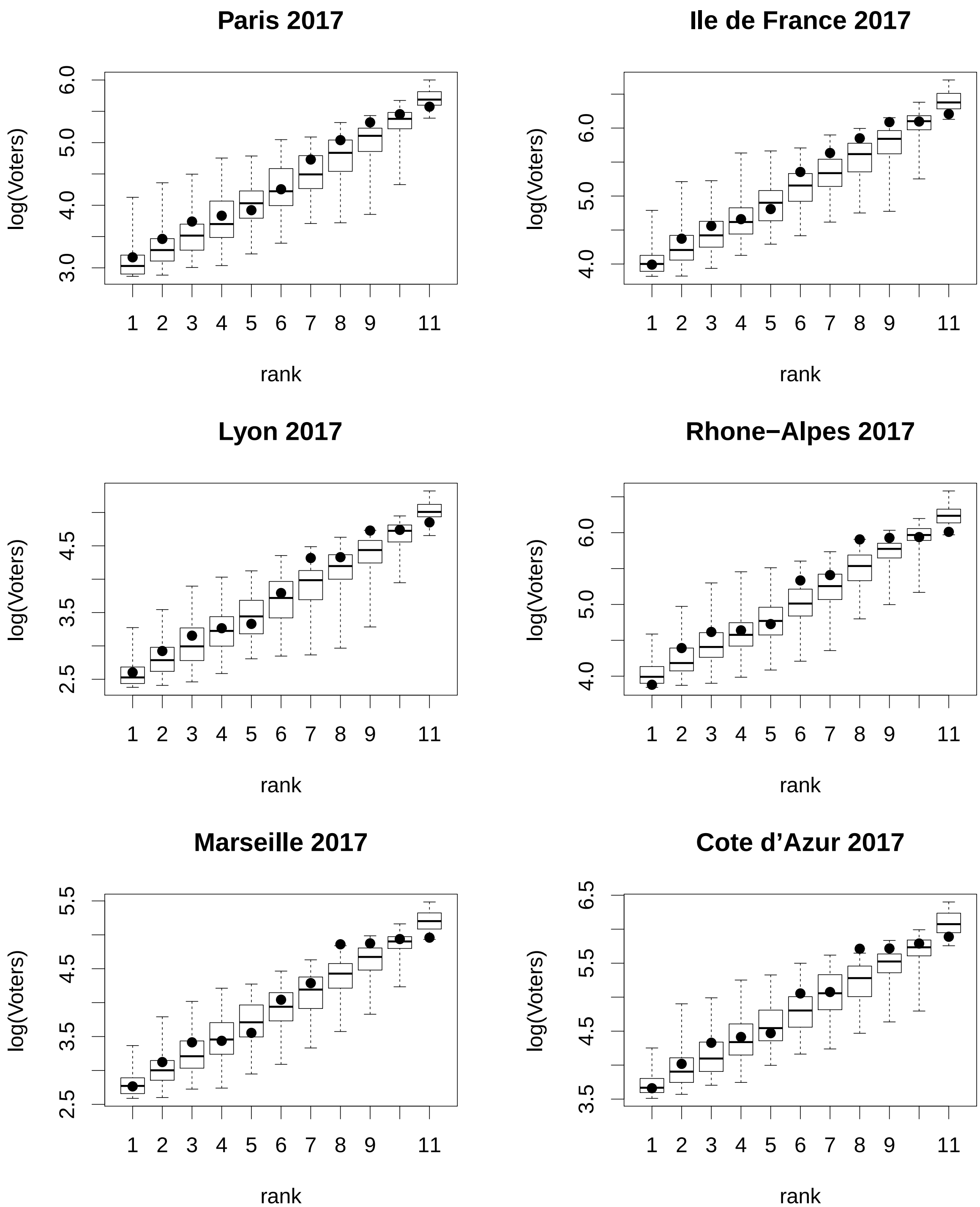


Figure 36: Election France, 2017 (boxplot of 100 realizations og the model, bullets: data).

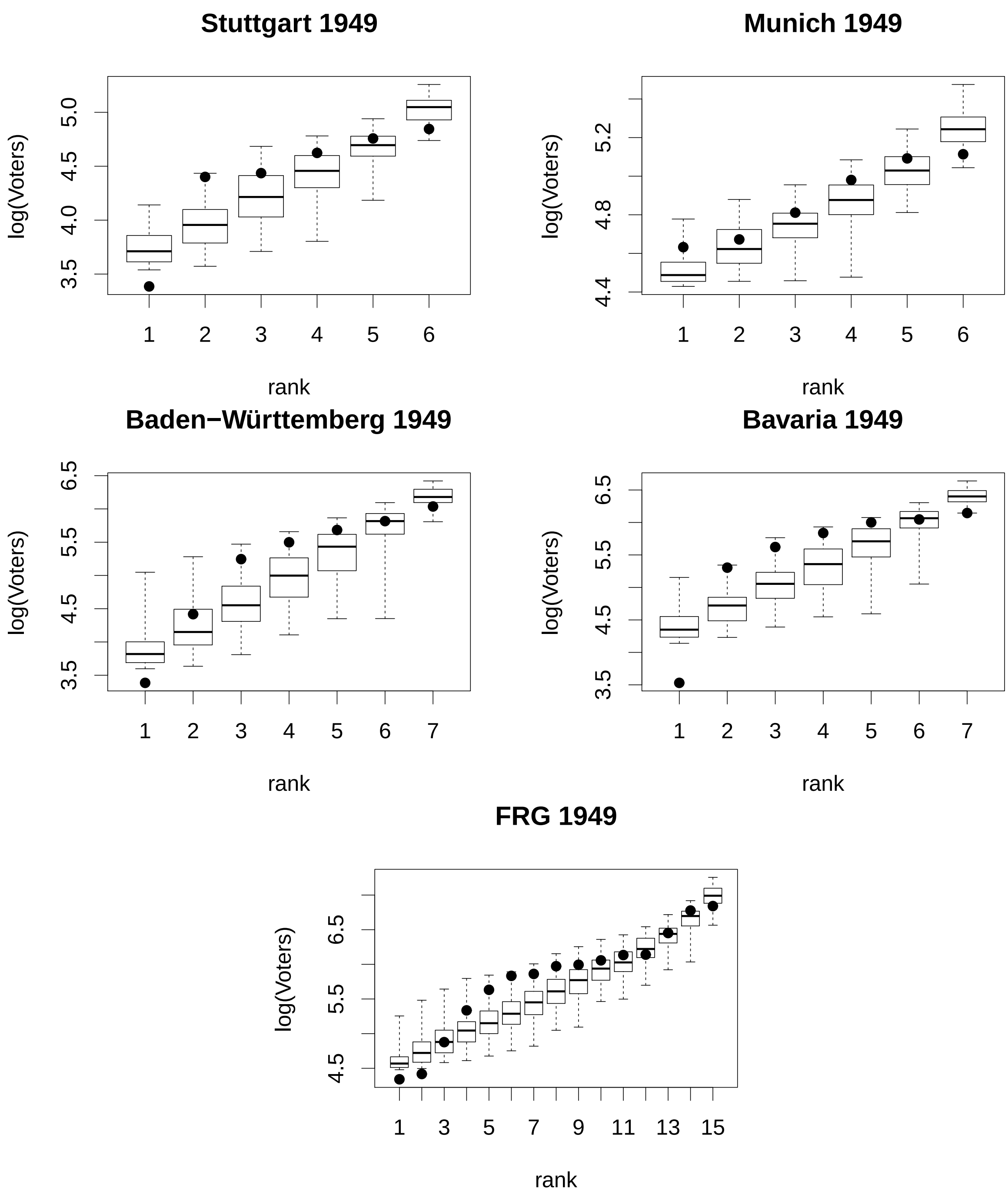


Figure 37: Election FRG, 1949 (boxplot of 100 realizations og the model, bullets: data).

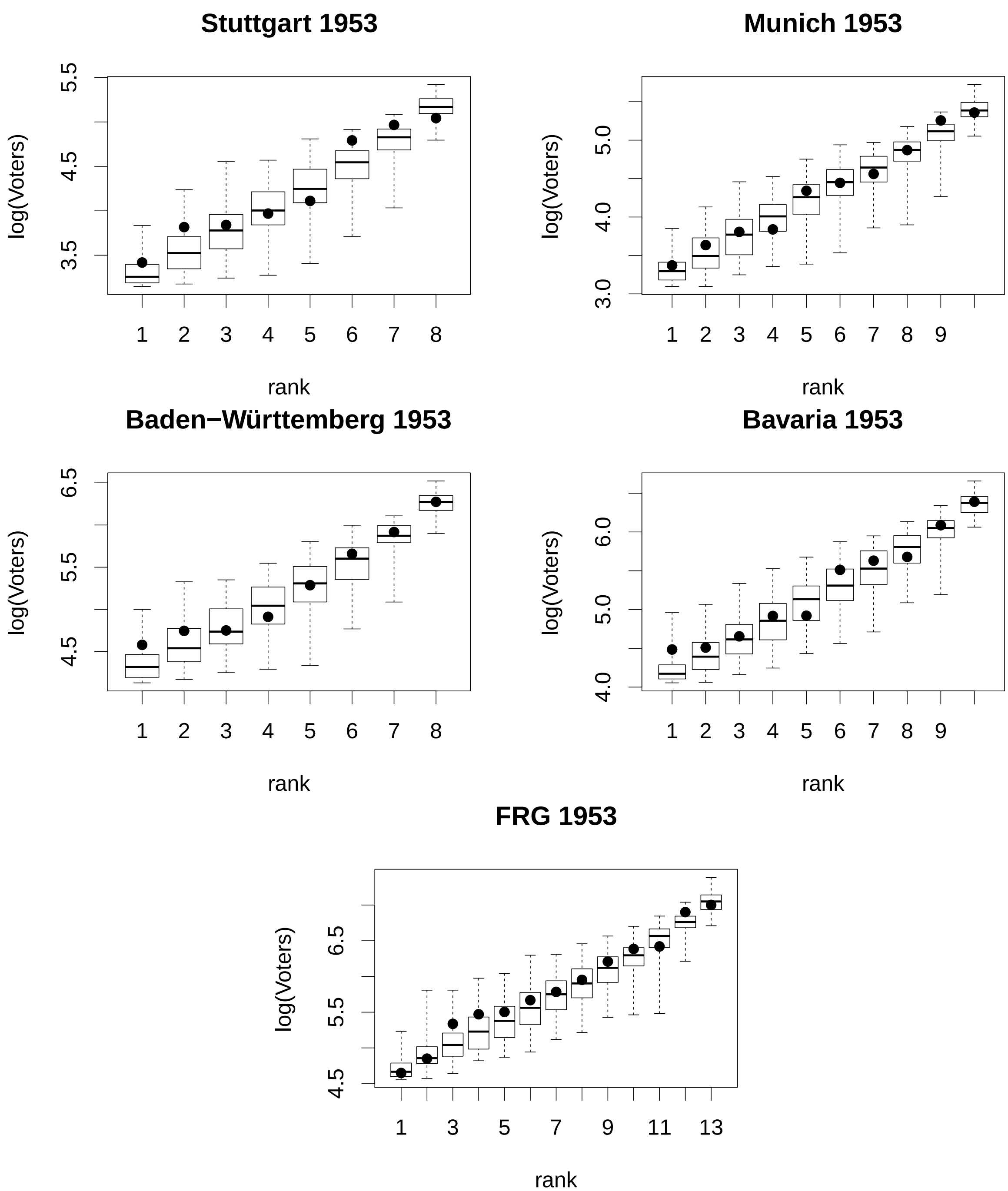


Figure 38: Election FRG, 1953 (boxplot of 100 realizations og the model, bullets: data).

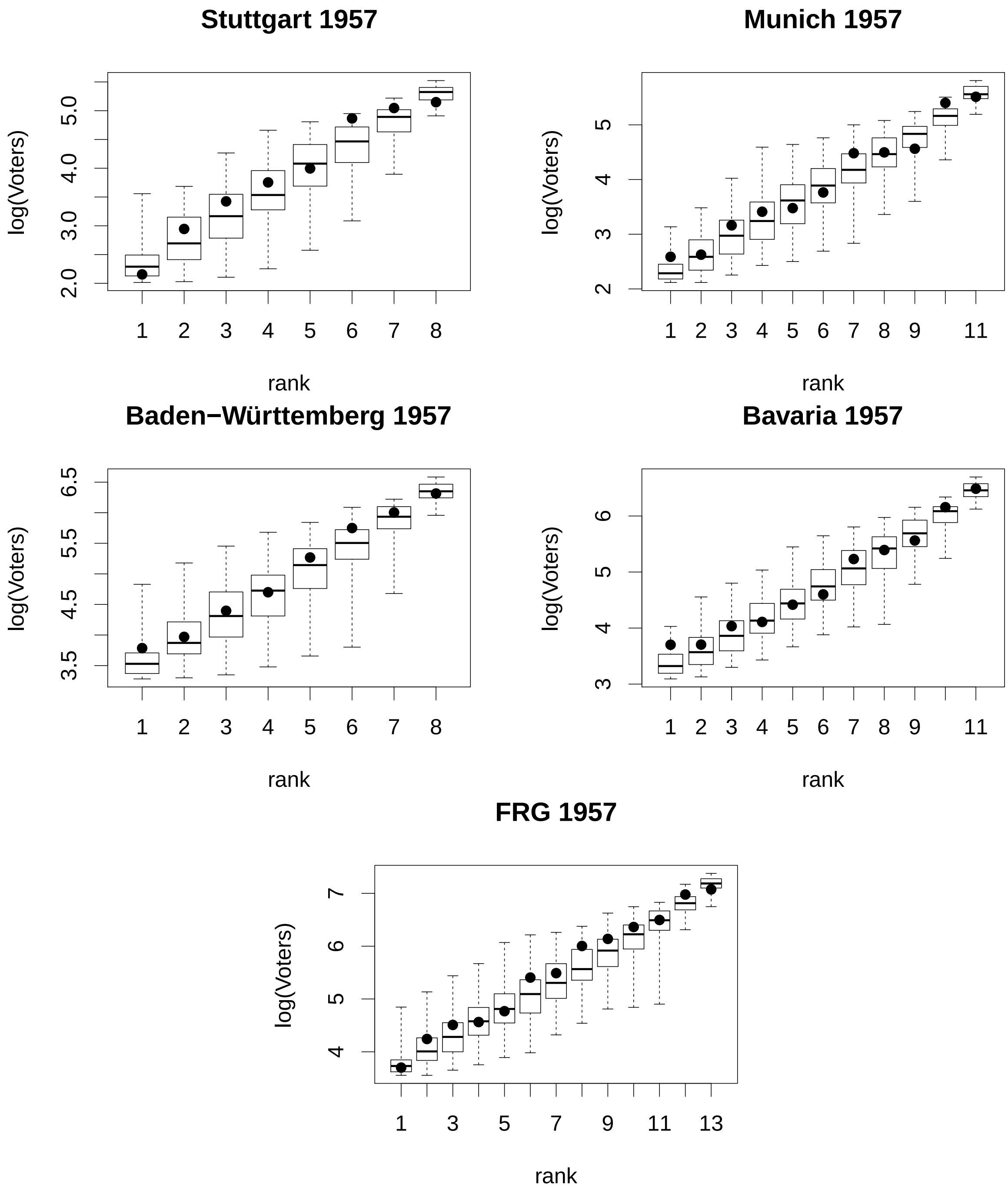


Figure 39: Election FRG, 1957 (boxplot of 100 realizations og the model, bullets: data).

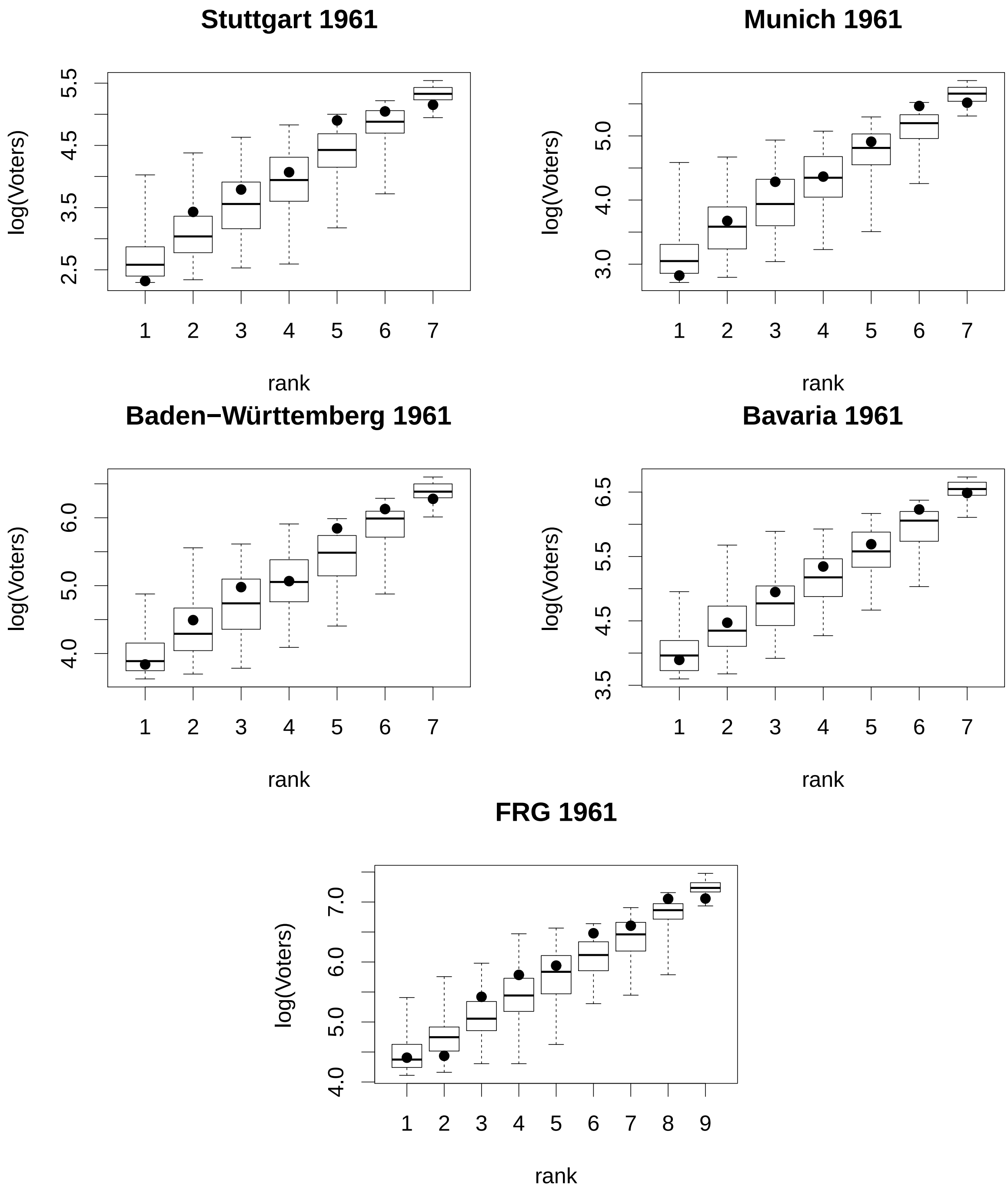


Figure 40: Election FRG, 1961 (boxplot of 100 realizations og the model, bullets: data).

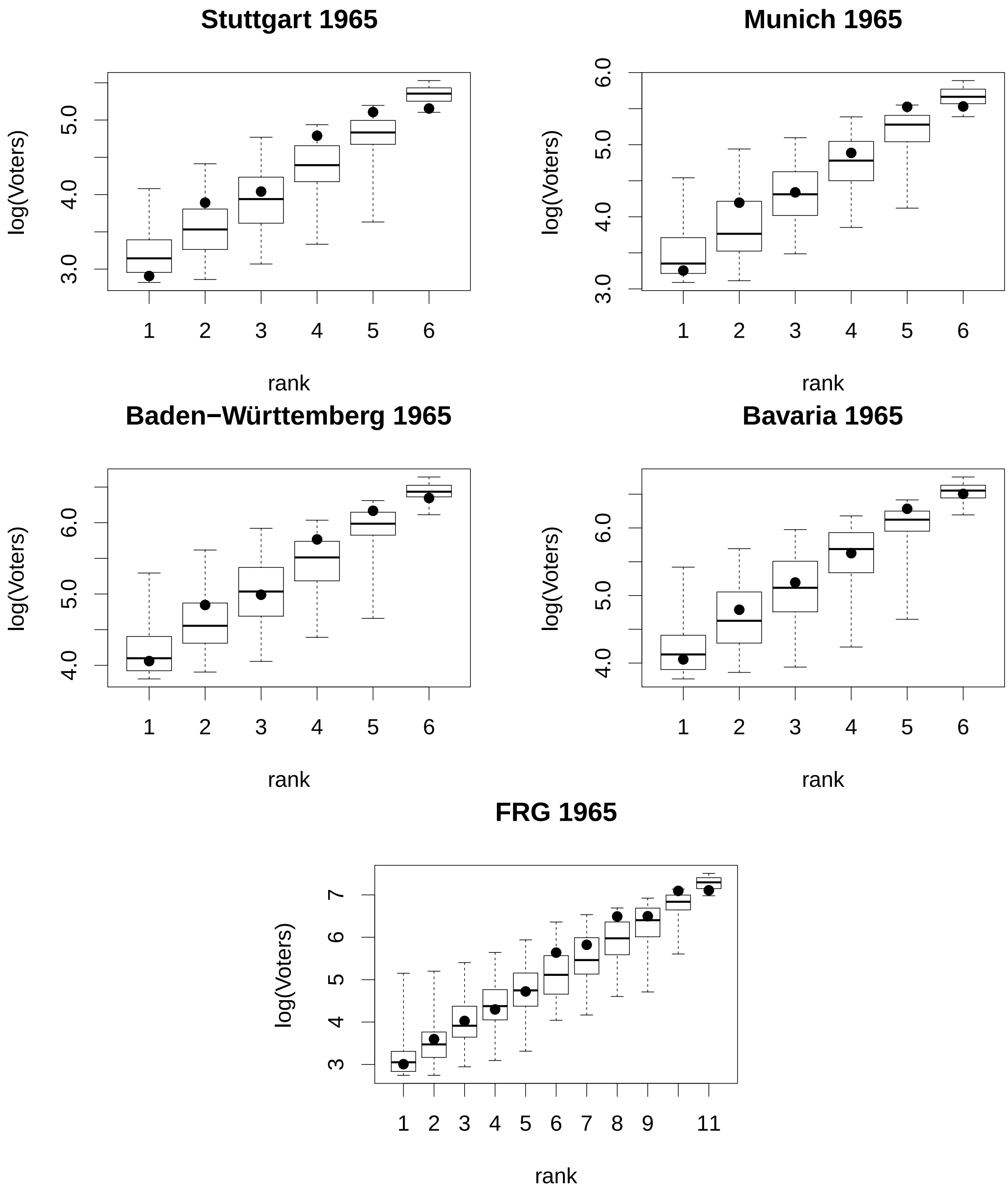


Figure 41: Election FRG, 1965 (boxplot of 100 realizations og the model, bullets: data).

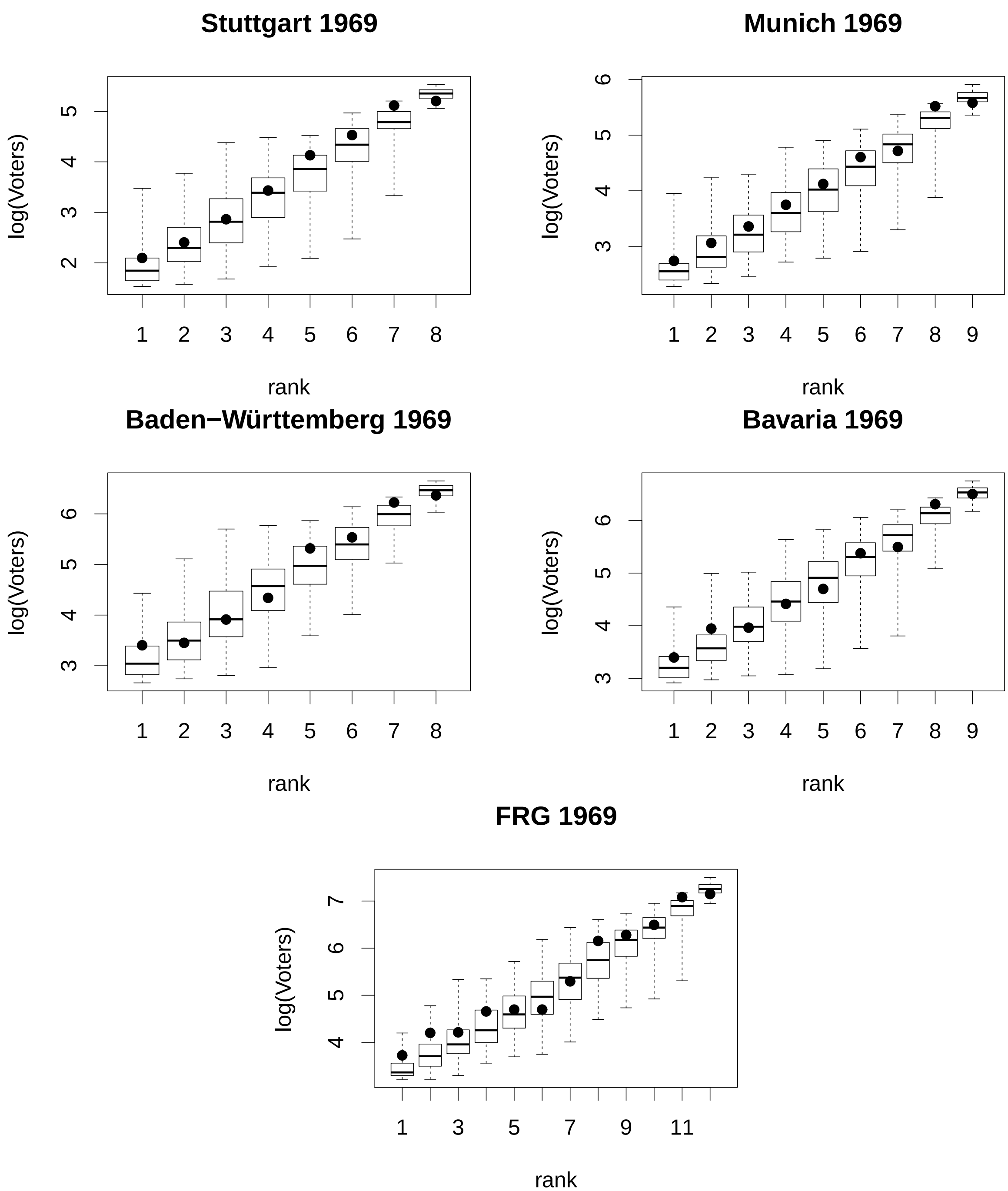


Figure 42: Election FRG, 1969 (boxplot of 100 realizations og the model, bullets: data).

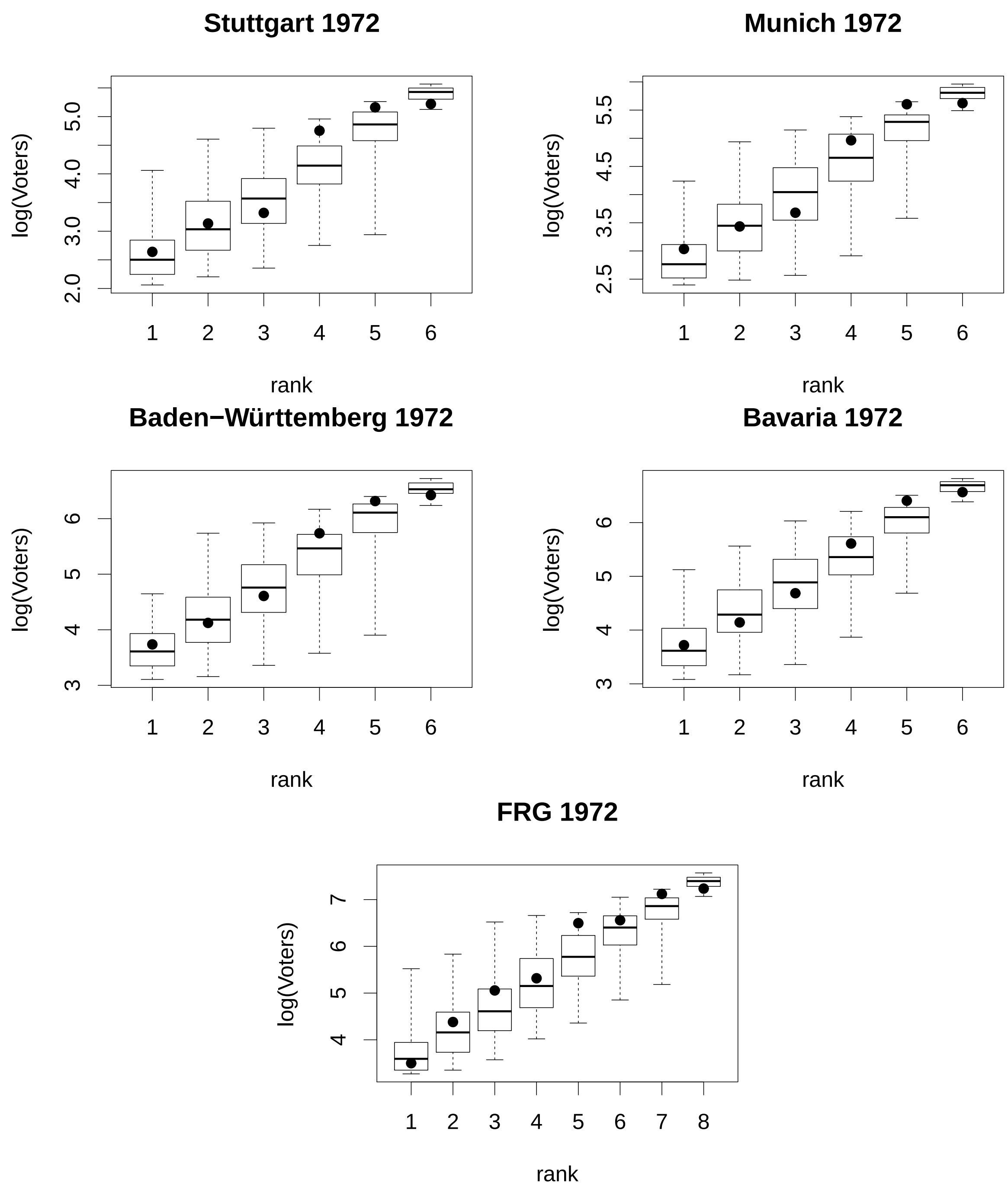


Figure 43: Election FRG, 1972 (boxplot of 100 realizations og the model, bullets: data).

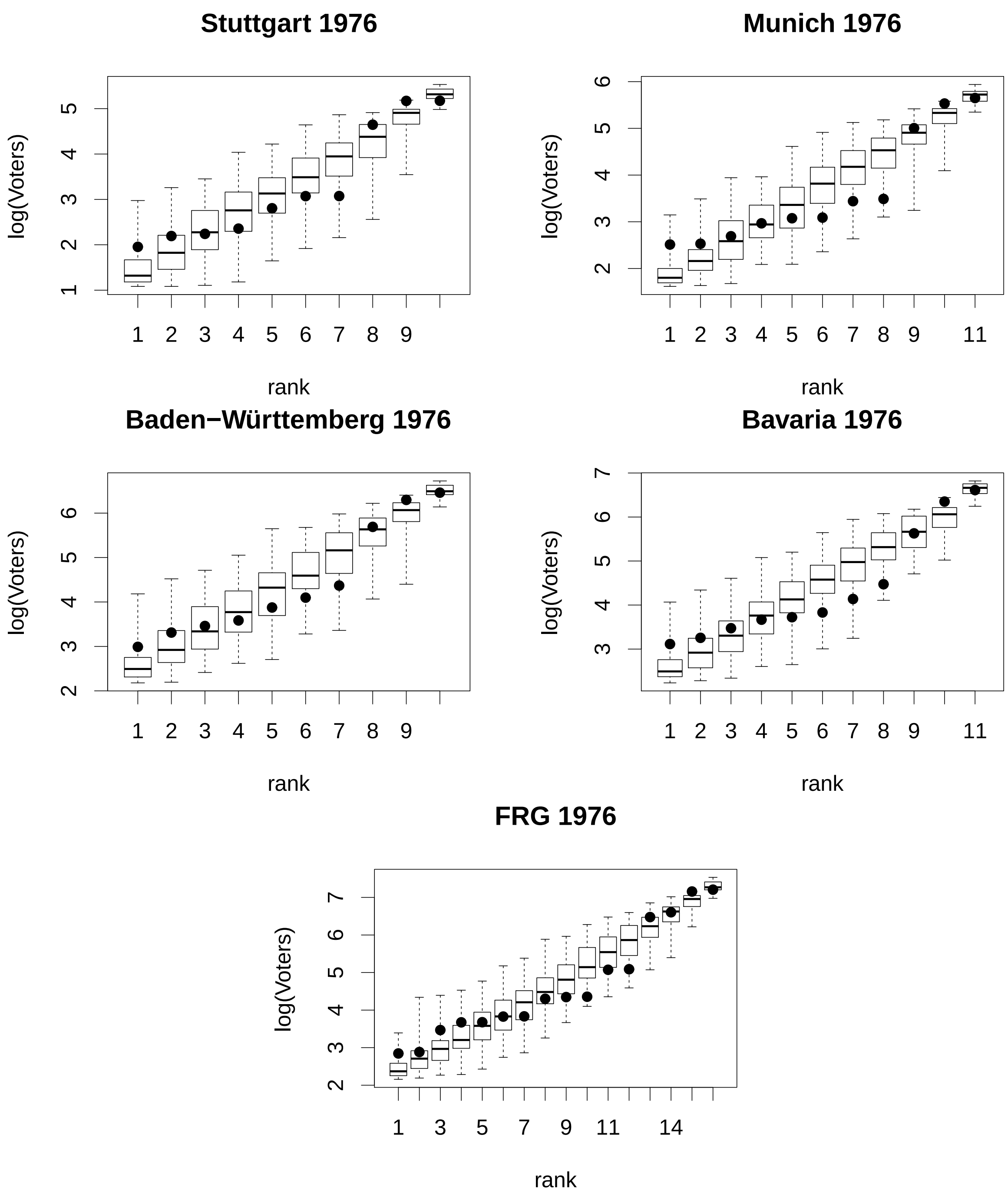


Figure 44: Election FRG, 1976 (boxplot of 100 realizations og the model, bullets: data).

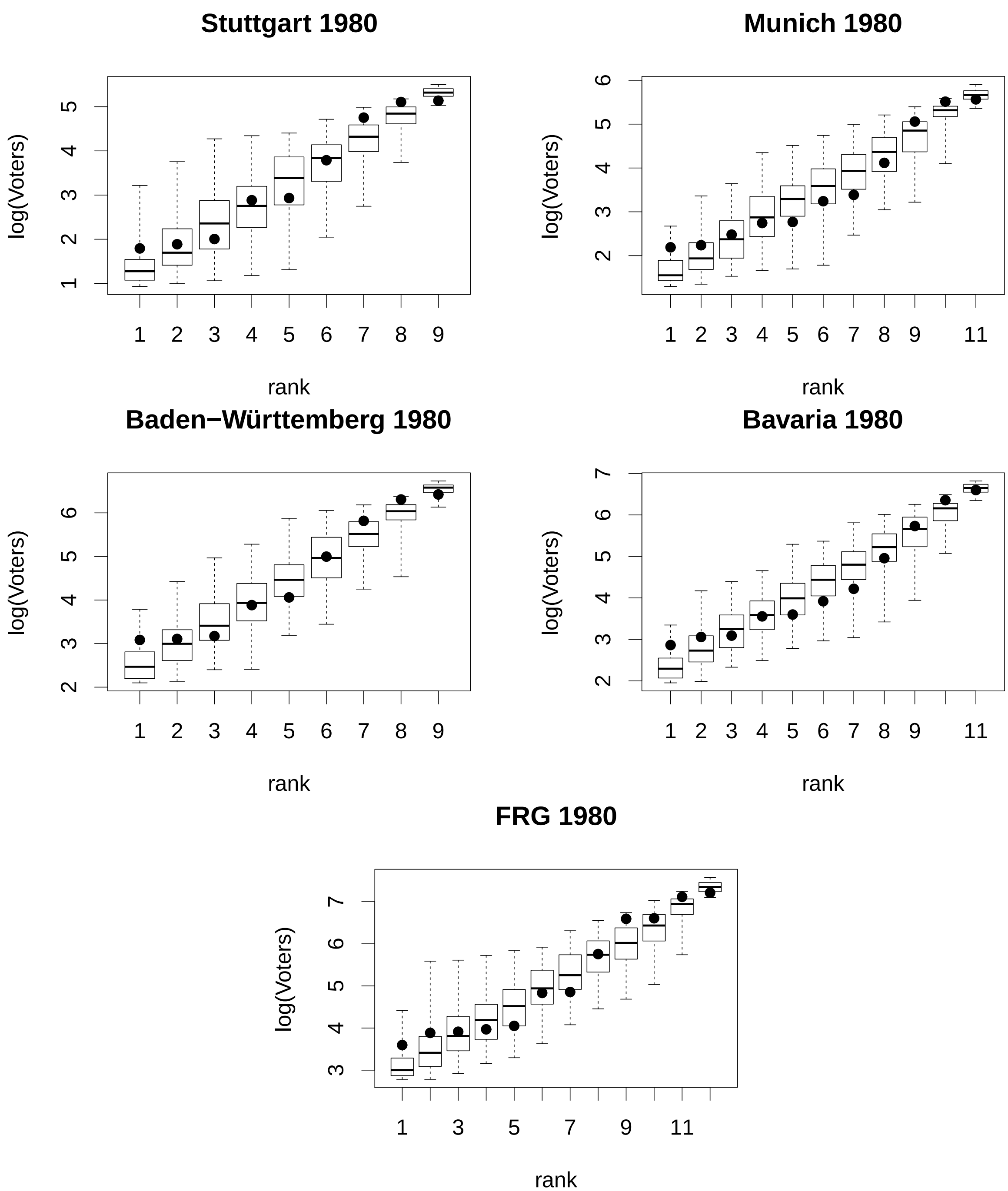


Figure 45: Election FRG, 1980 (boxplot of 100 realizations og the model, bullets: data).

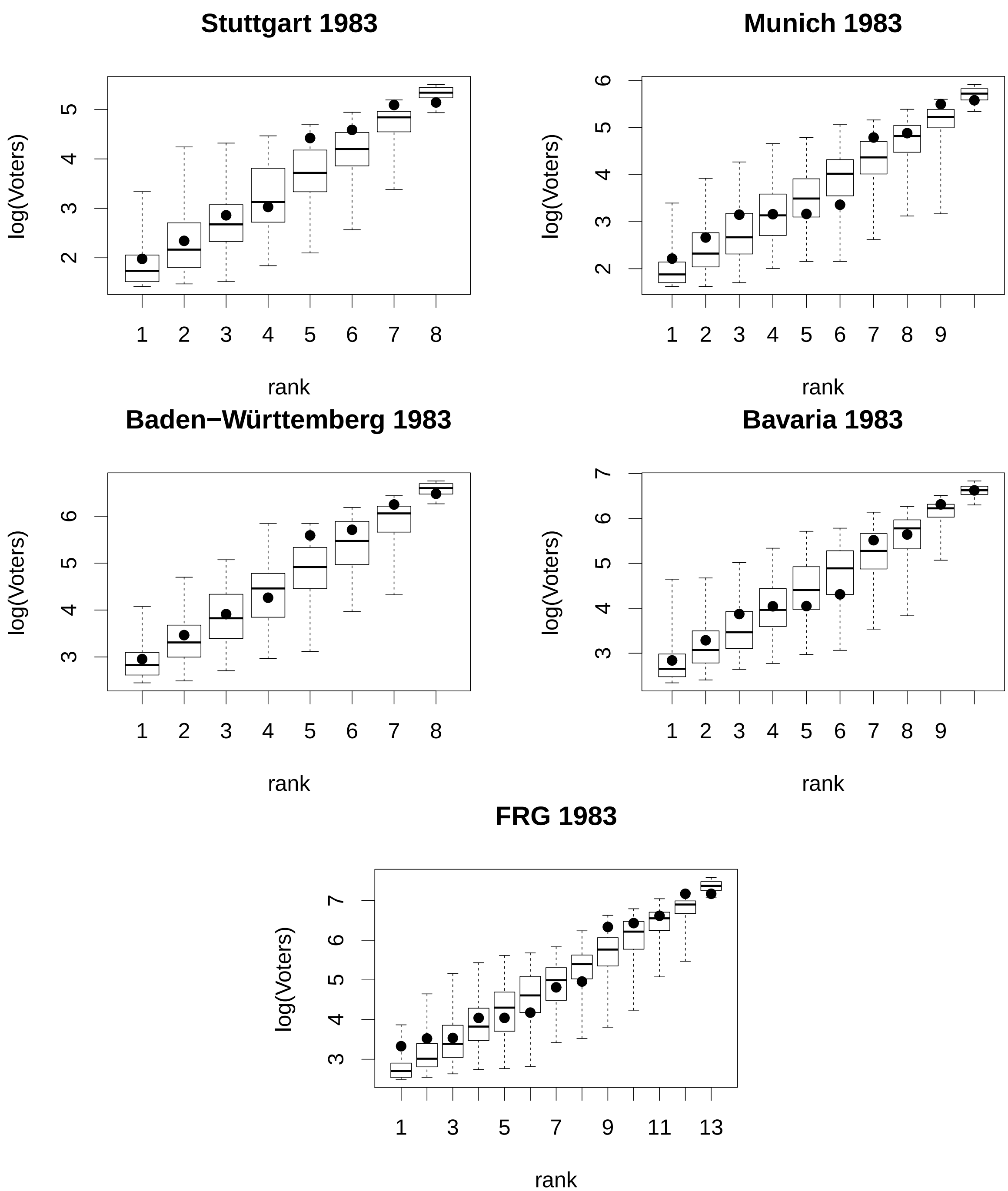


Figure 46: Election FRG, 1983 (boxplot of 100 realizations og the model, bullets: data).

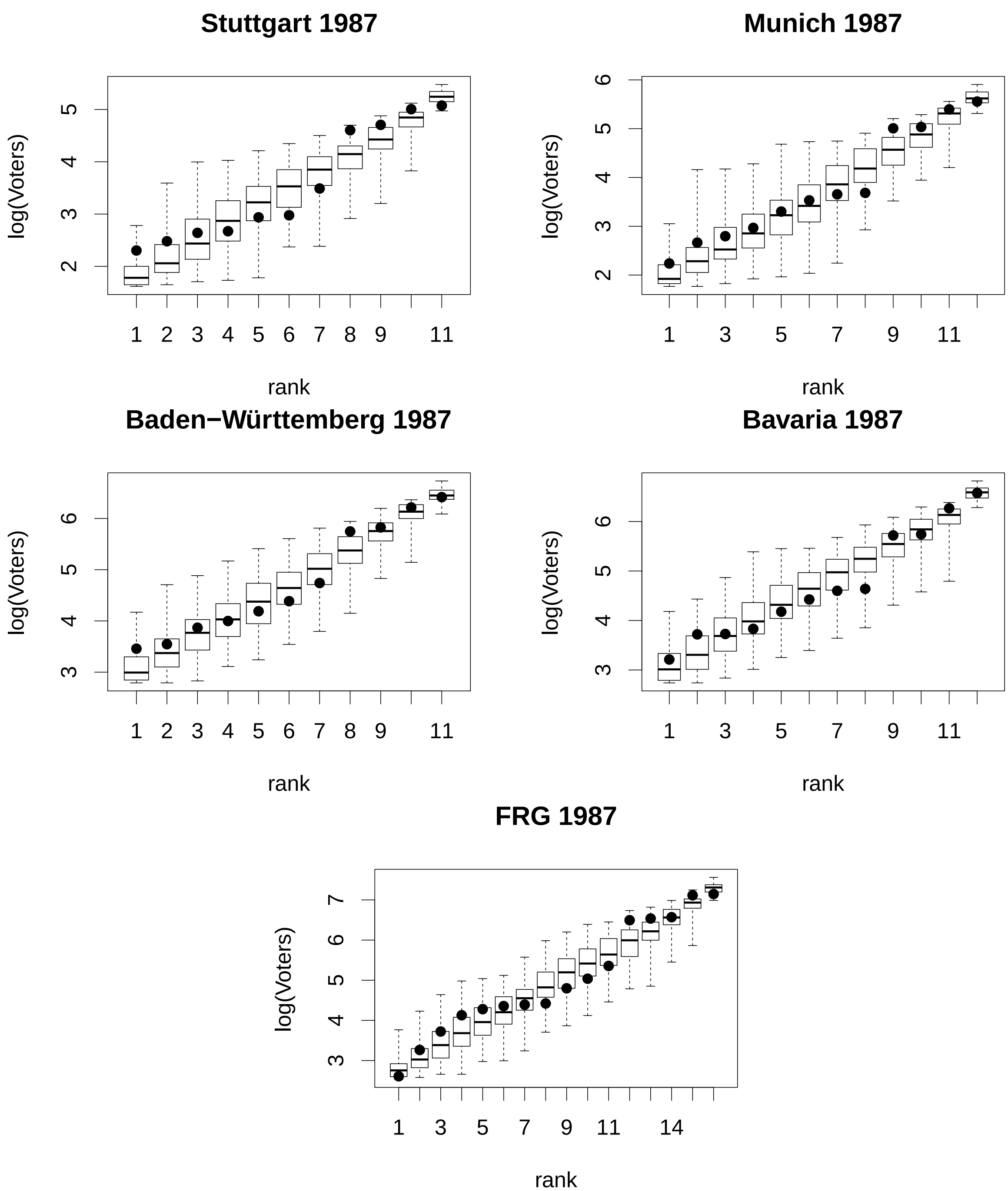


Figure 47: Election FRG, 1987 (boxplot of 100 realizations og the model, bullets: data).

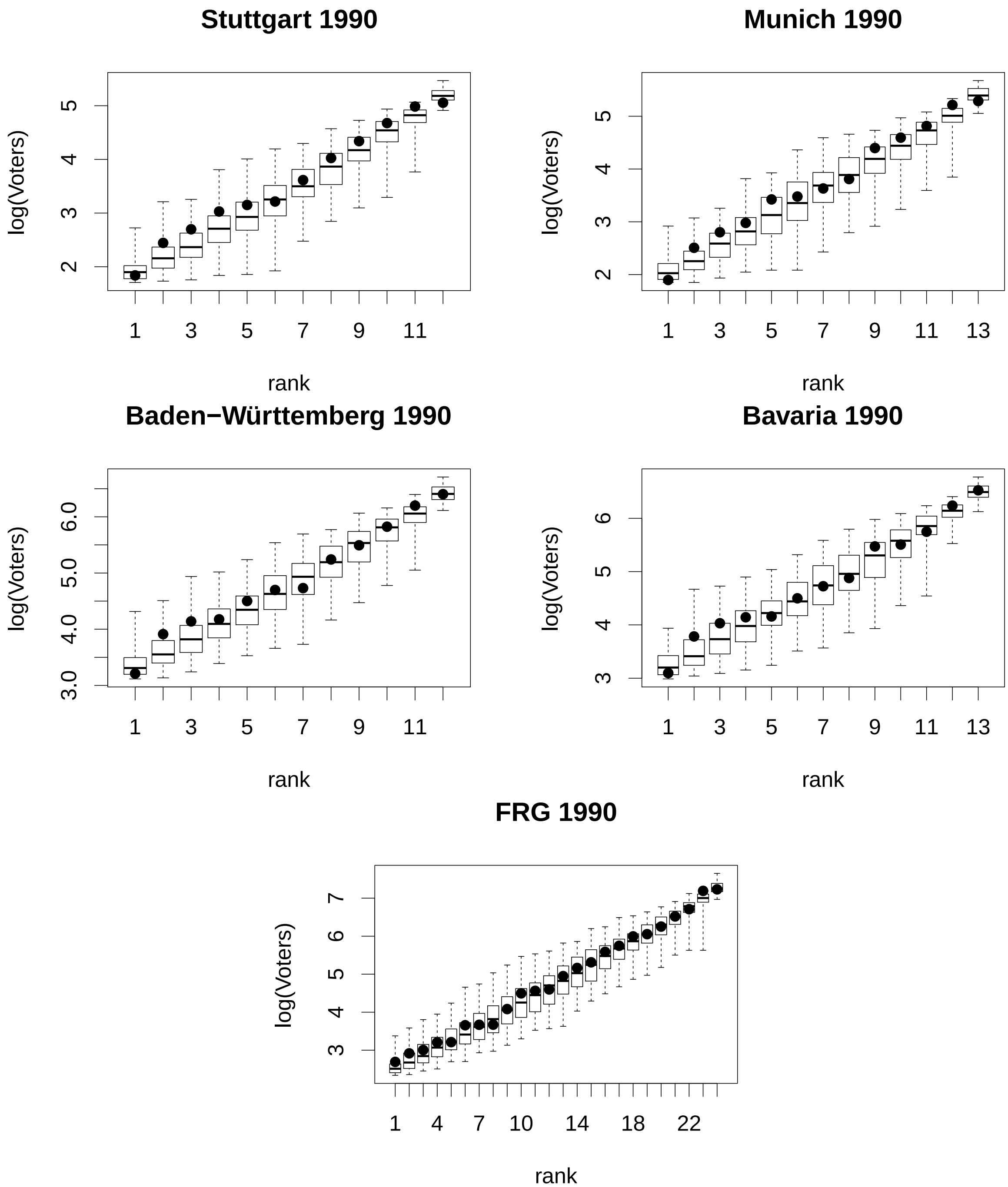


Figure 48: Election FRG, 1990 (boxplot of 100 realizations og the model, bullets: data).

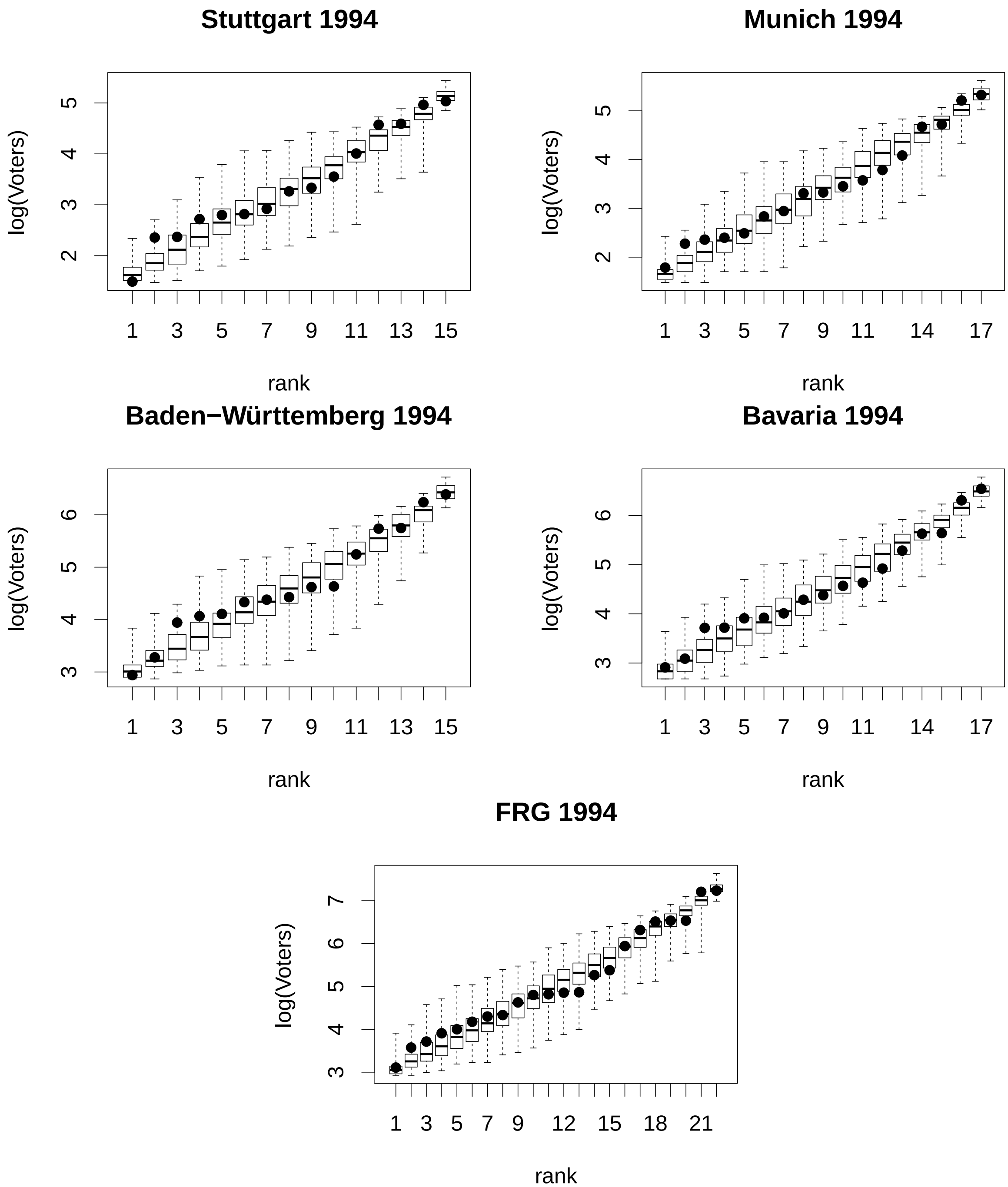


Figure 49: Election FRG, 1994 (boxplot of 100 realizations og the model, bullets: data).

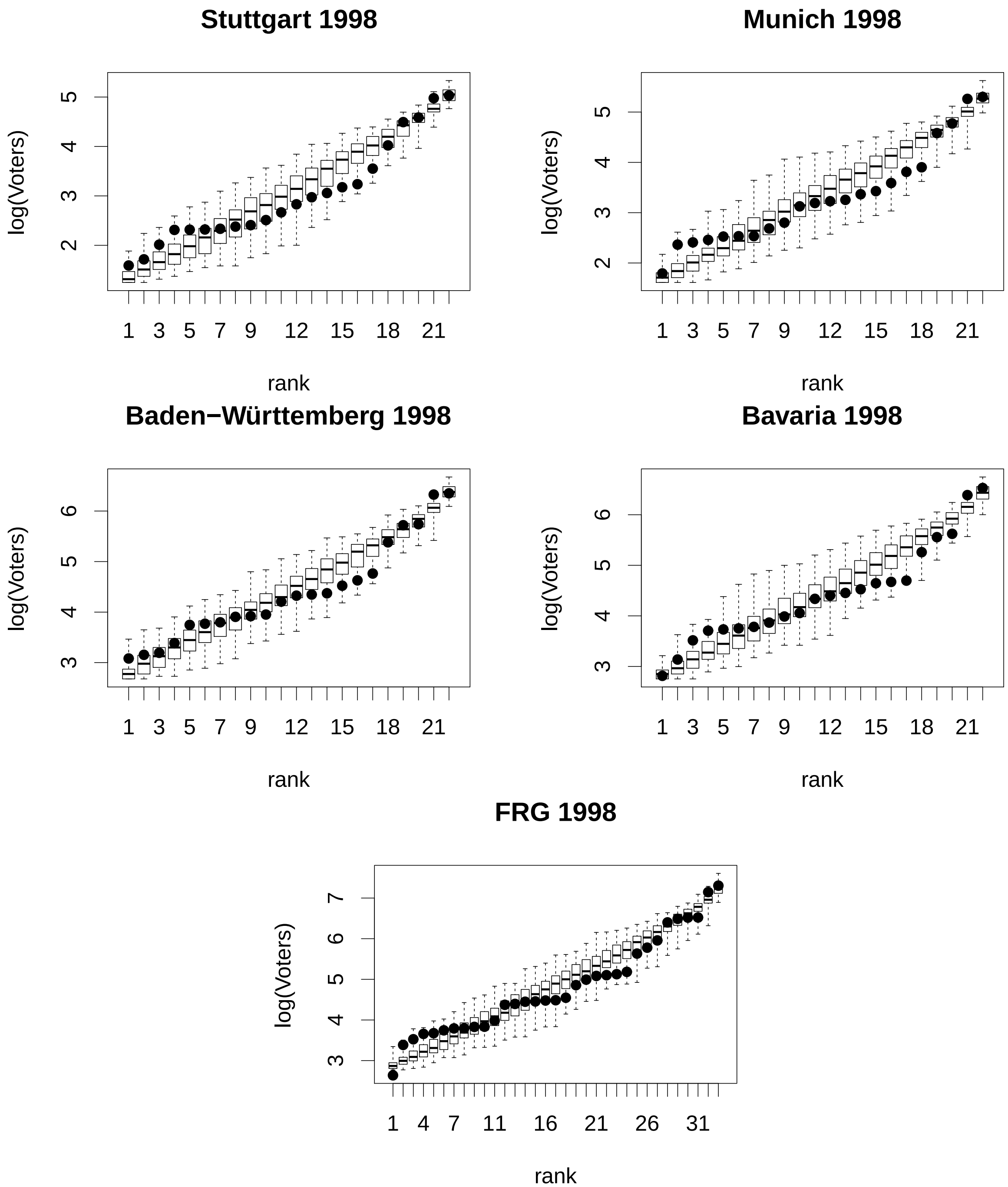


Figure 50: Election FRG, 1998 (boxplot of 100 realizations og the model, bullets: data).

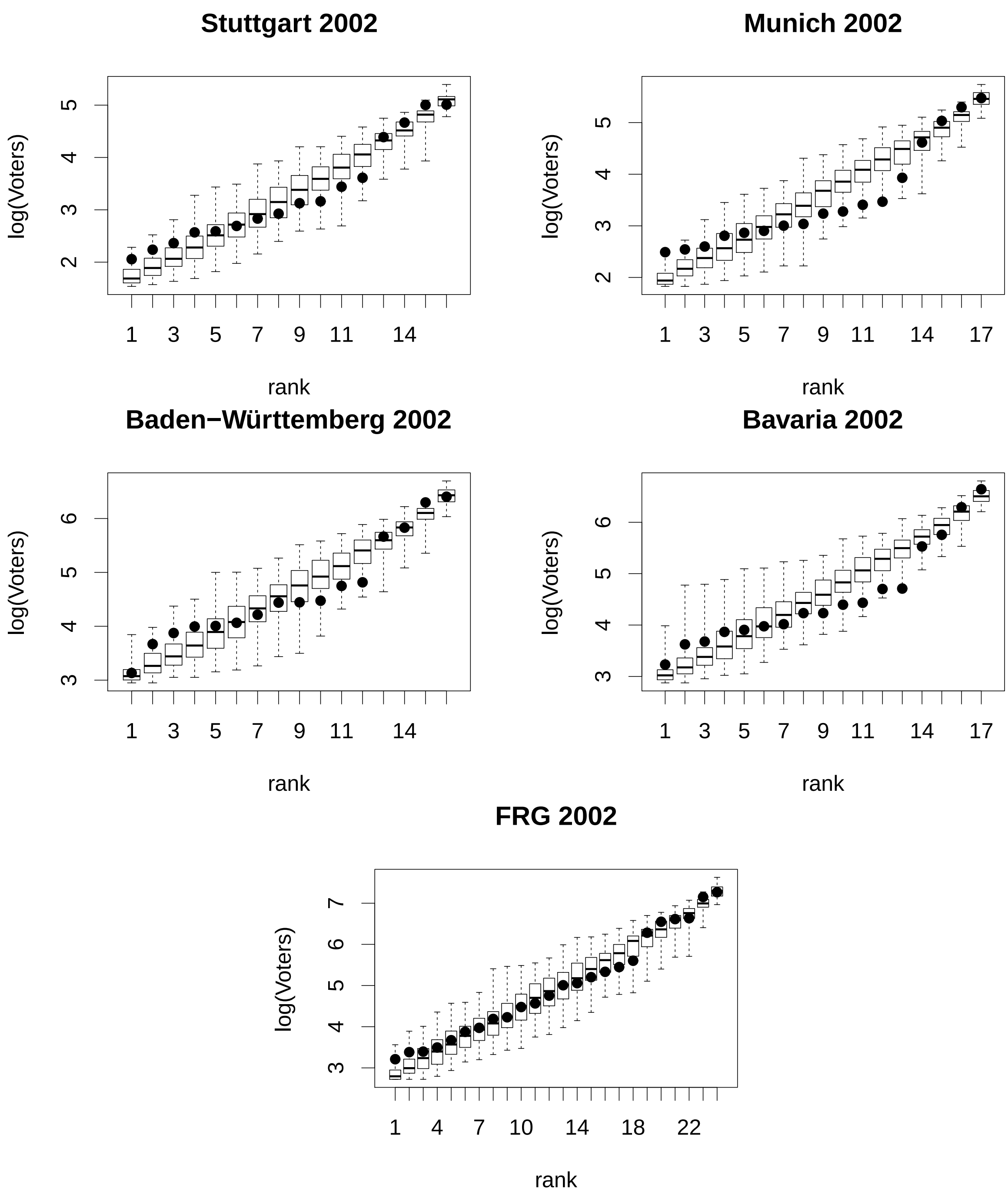


Figure 51: Election FRG, 2002 (boxplot of 100 realizations og the model, bullets: data).

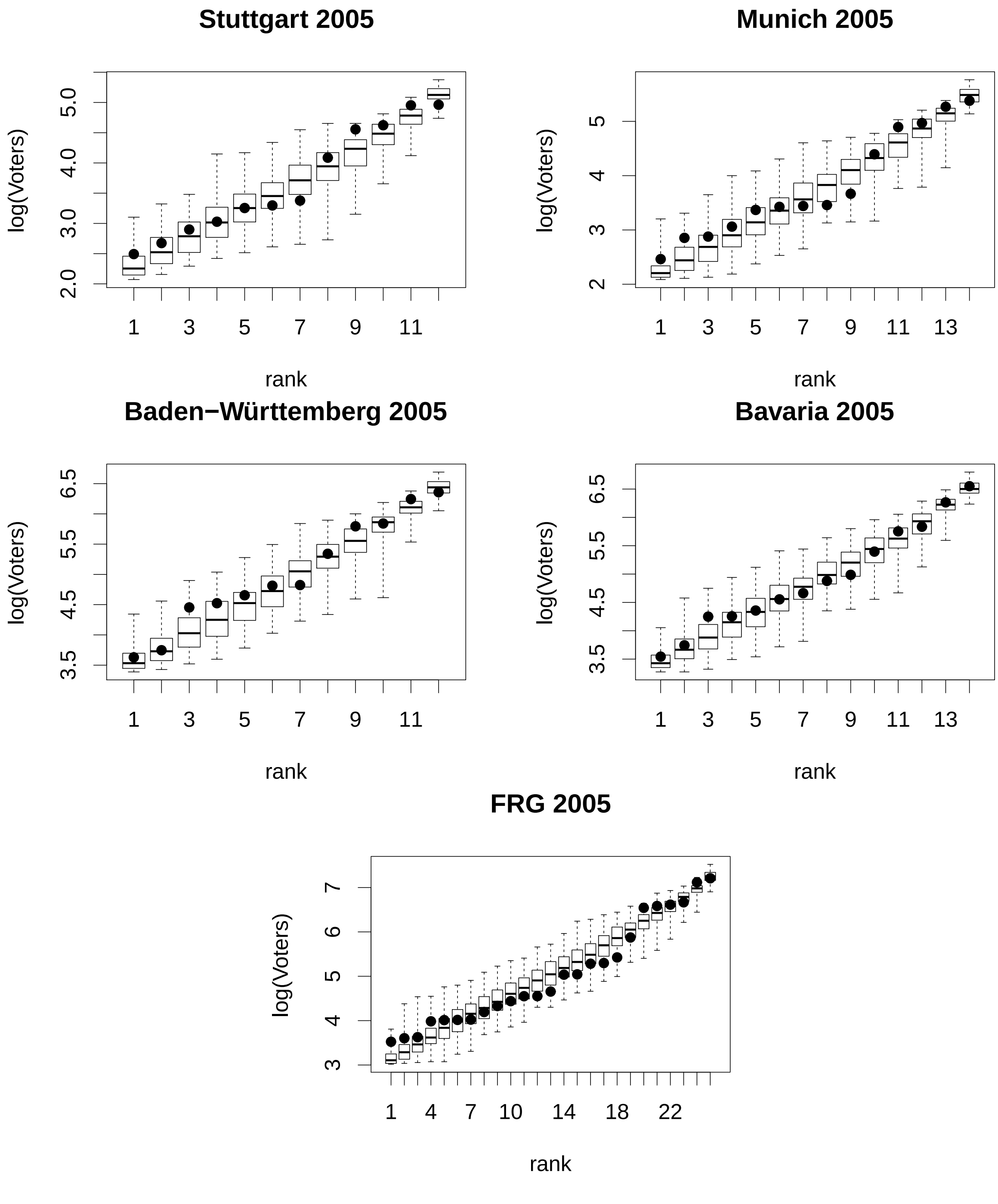


Figure 52: Election FRG, 2005 (boxplot of 100 realizations og the model, bullets: data).

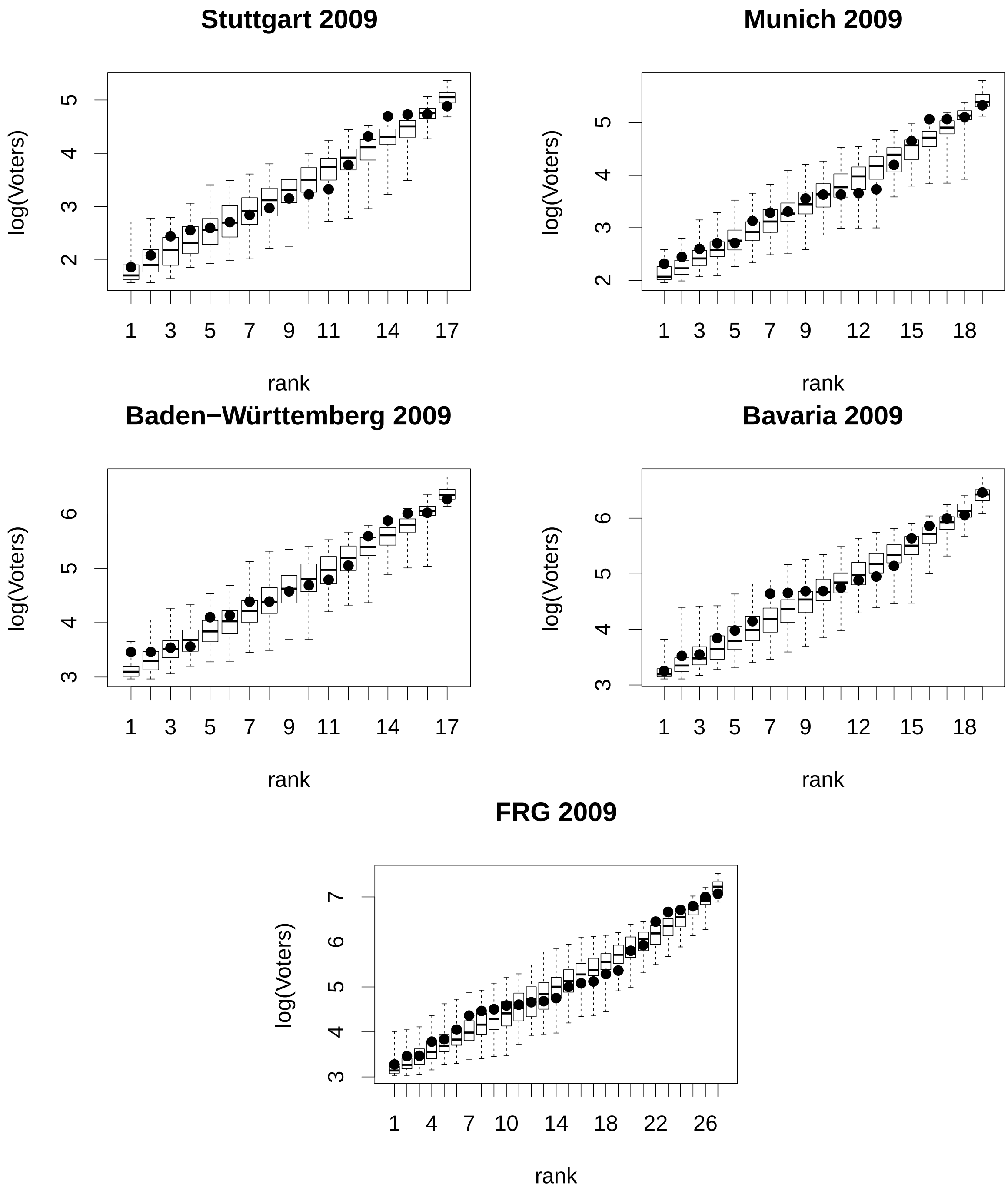


Figure 53: Election FRG, 2009 (boxplot of 100 realizations og the model, bullets: data).

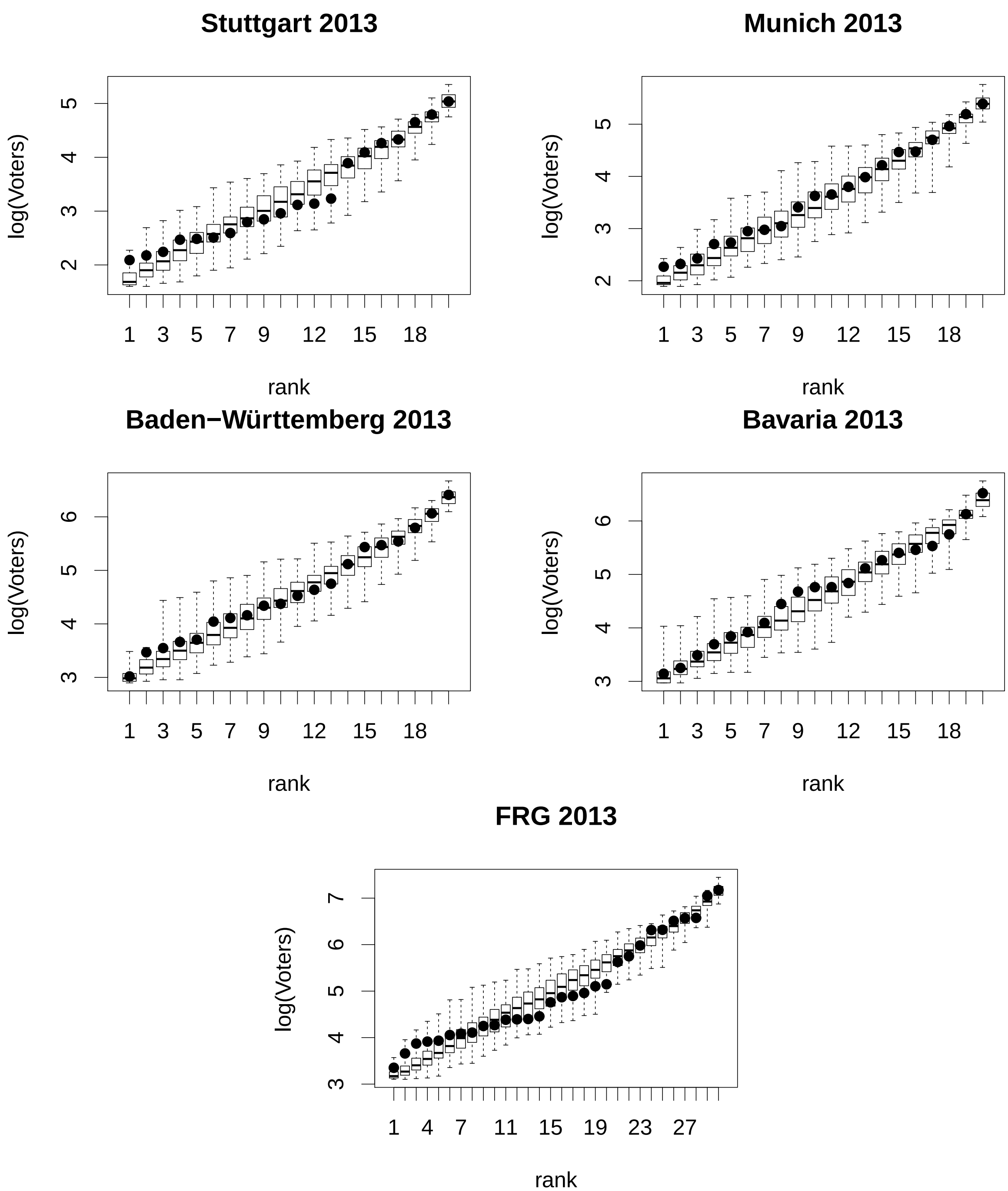


Figure 54: Election FRG, 2013 (boxplot of 100 realizations og the model, bullets: data).

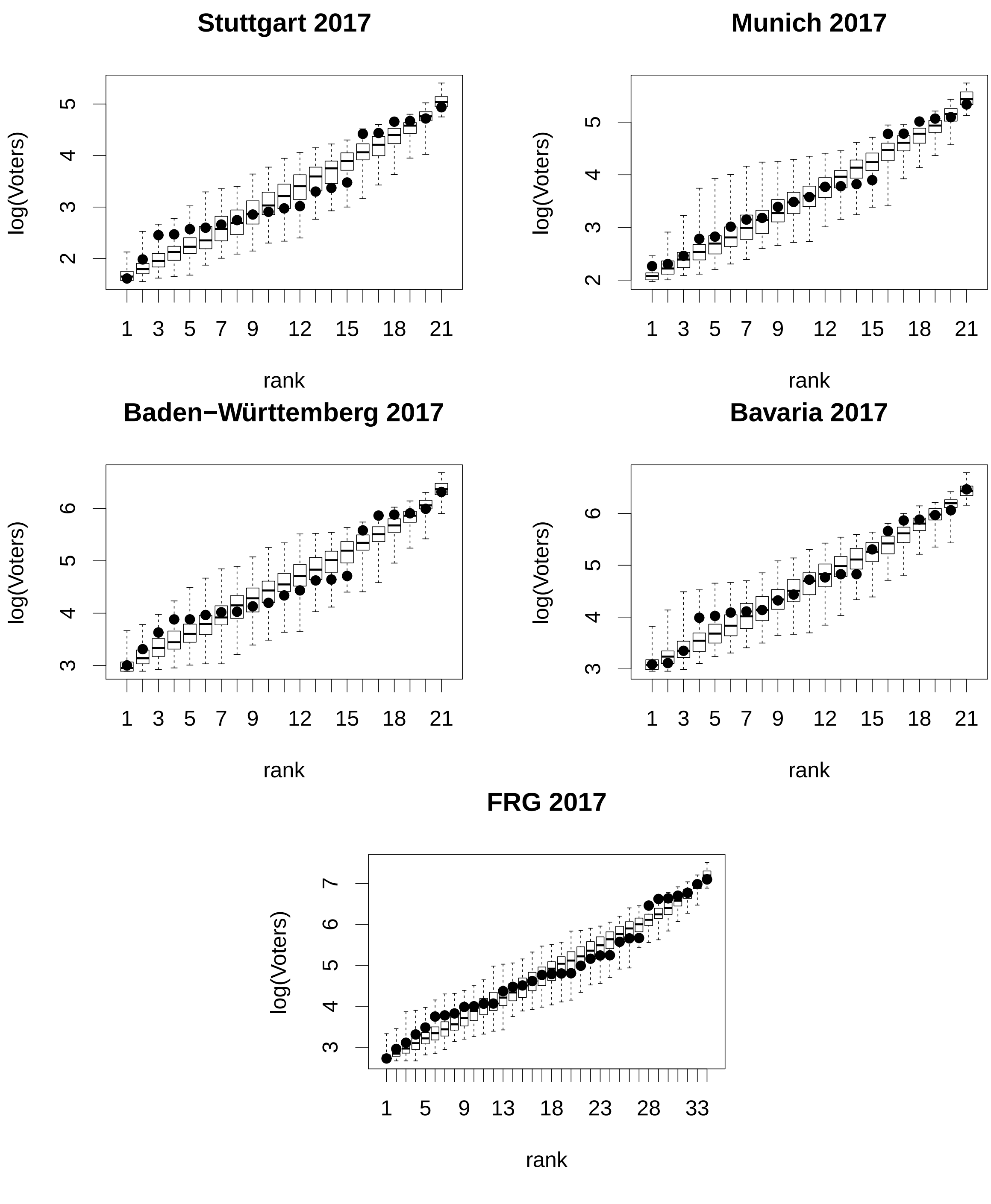


Figure 55: Election FRG, 2017 (boxplot of 100 realizations og the model, bullets: data).

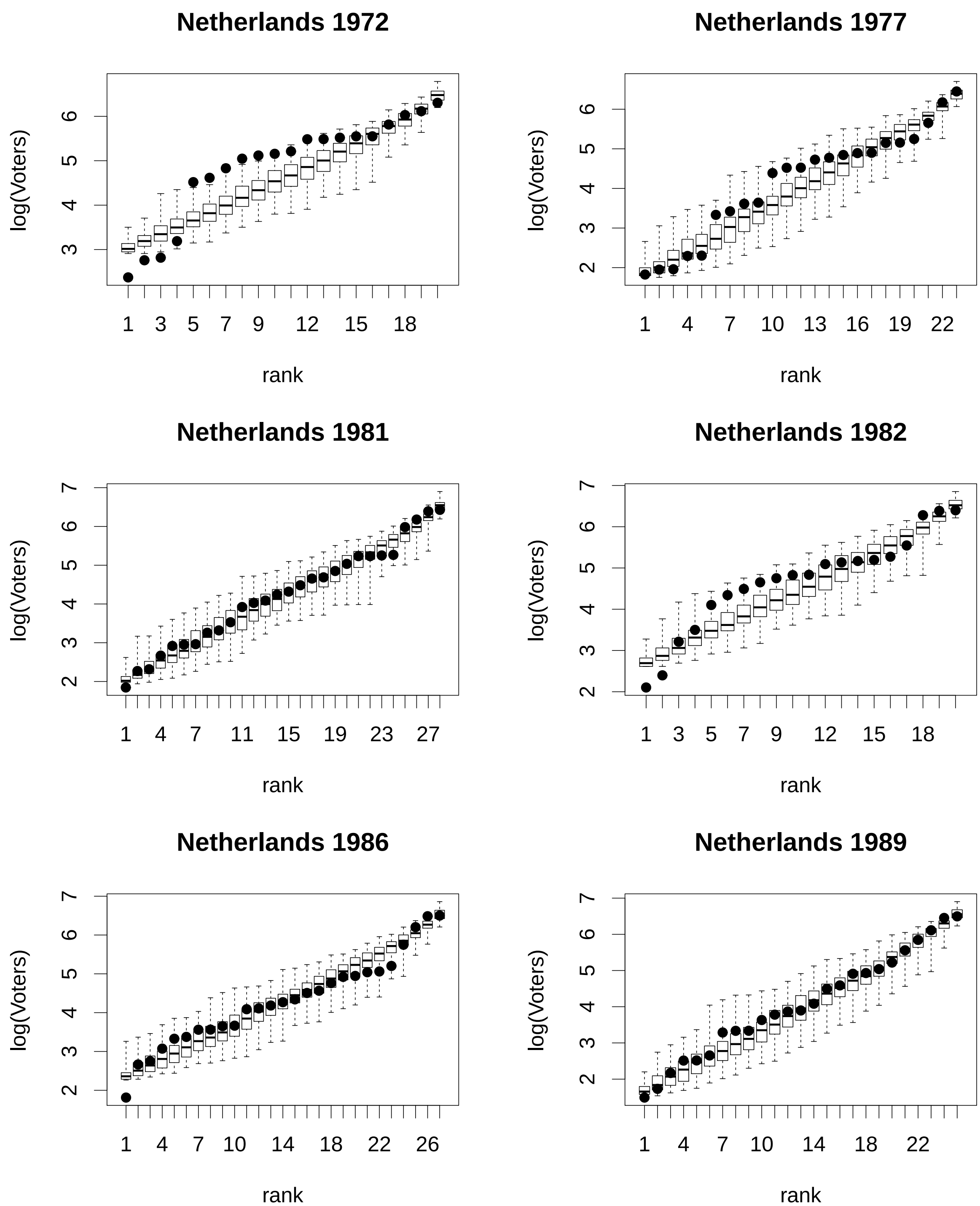


Figure 56: Election in the Netherlands (boxplot of 100 realizations og the model, bullets: data).

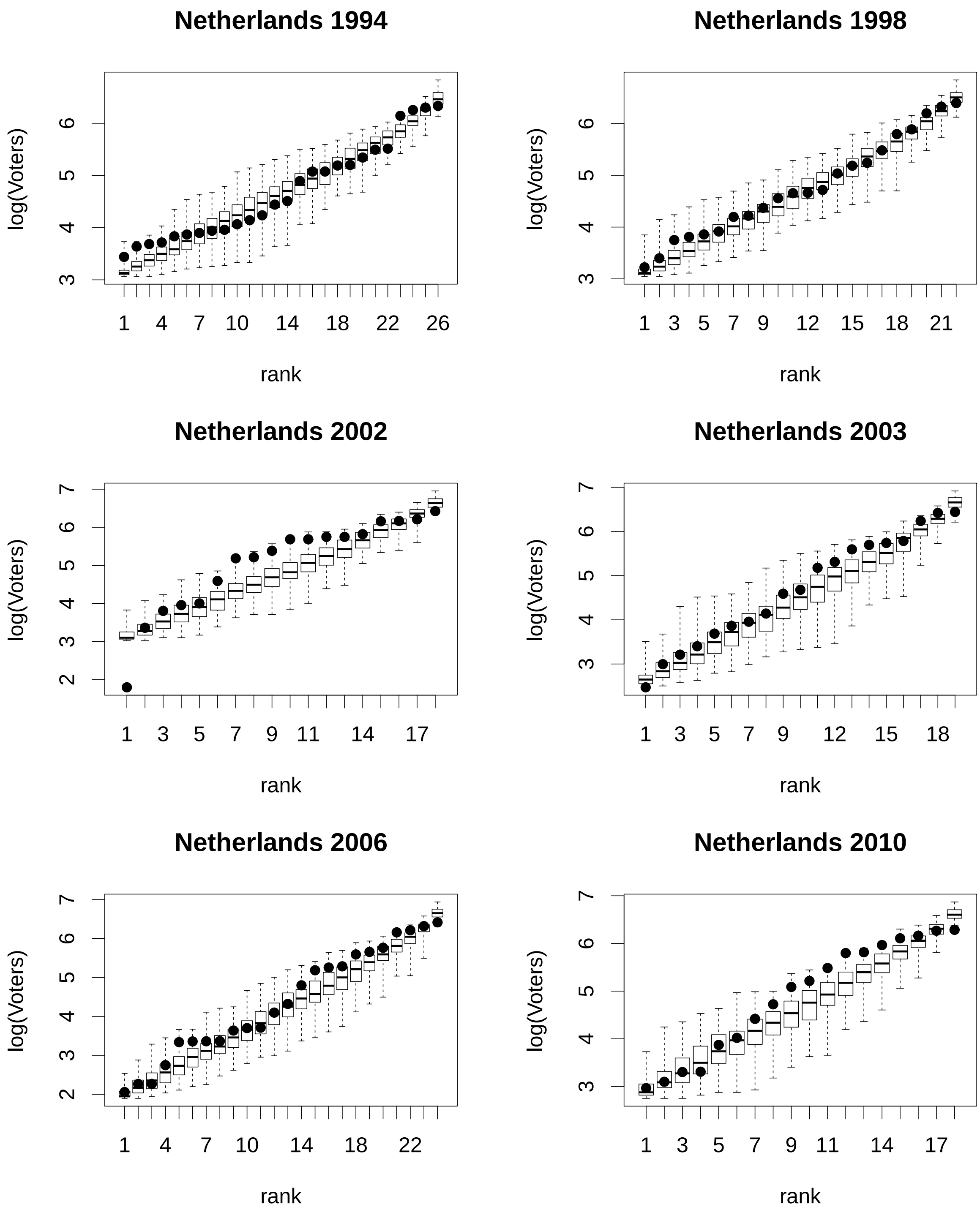


Figure 57: Election in the Netherlands (boxplot of 100 realizations og the model, bullets: data).

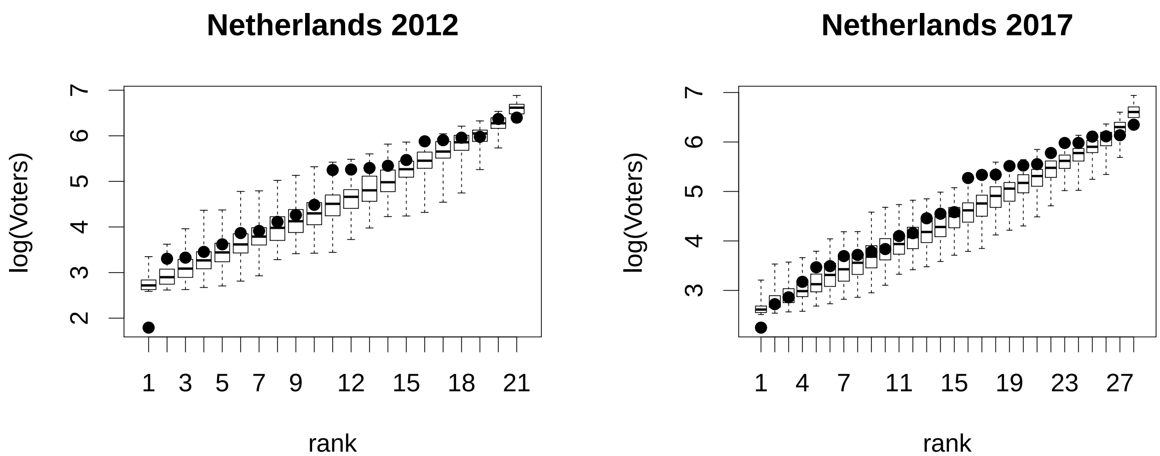


Figure 58: Election in the Netherlands (boxplot of 100 realizations og the model, bullets: data).

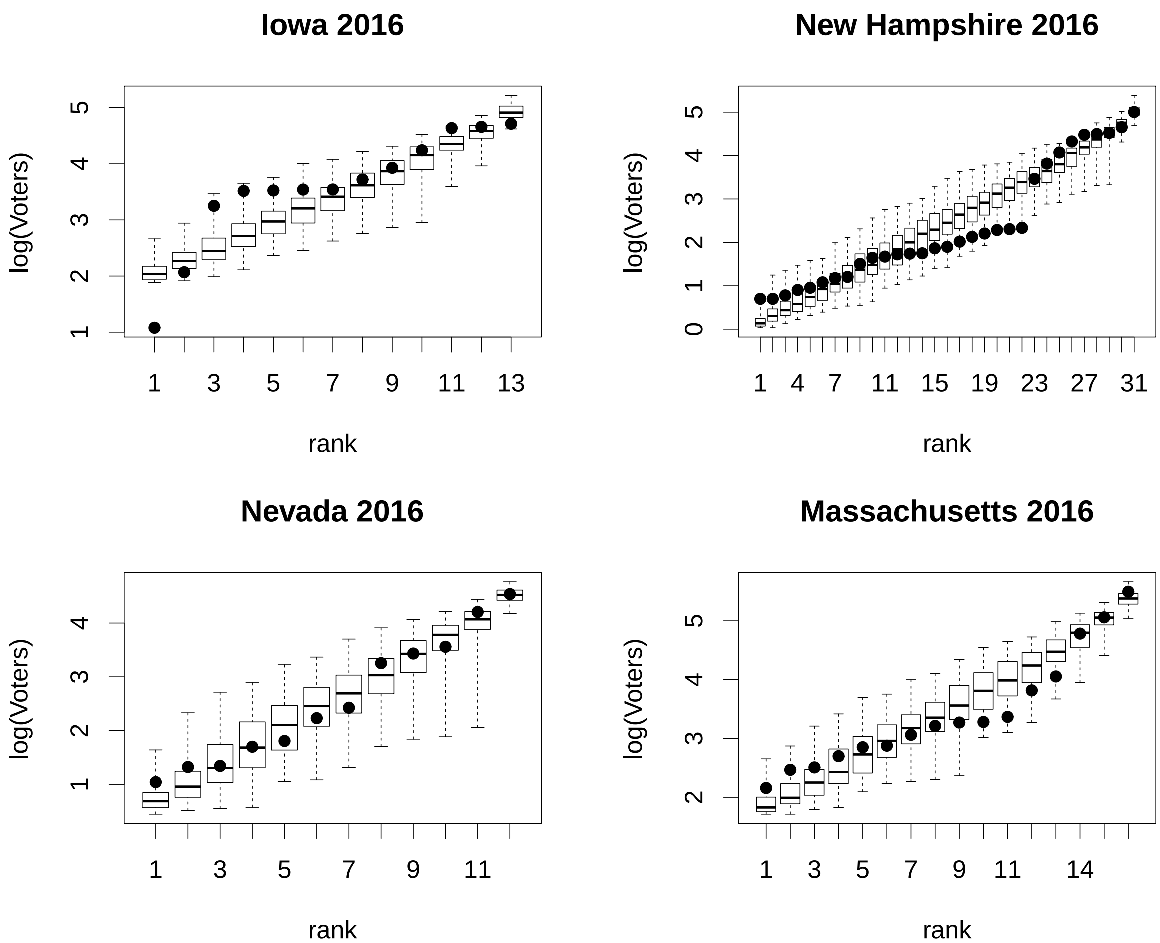


Figure 59: Election US (republicans), 2016 (boxplot of 100 realizations og the model, bullets: data).

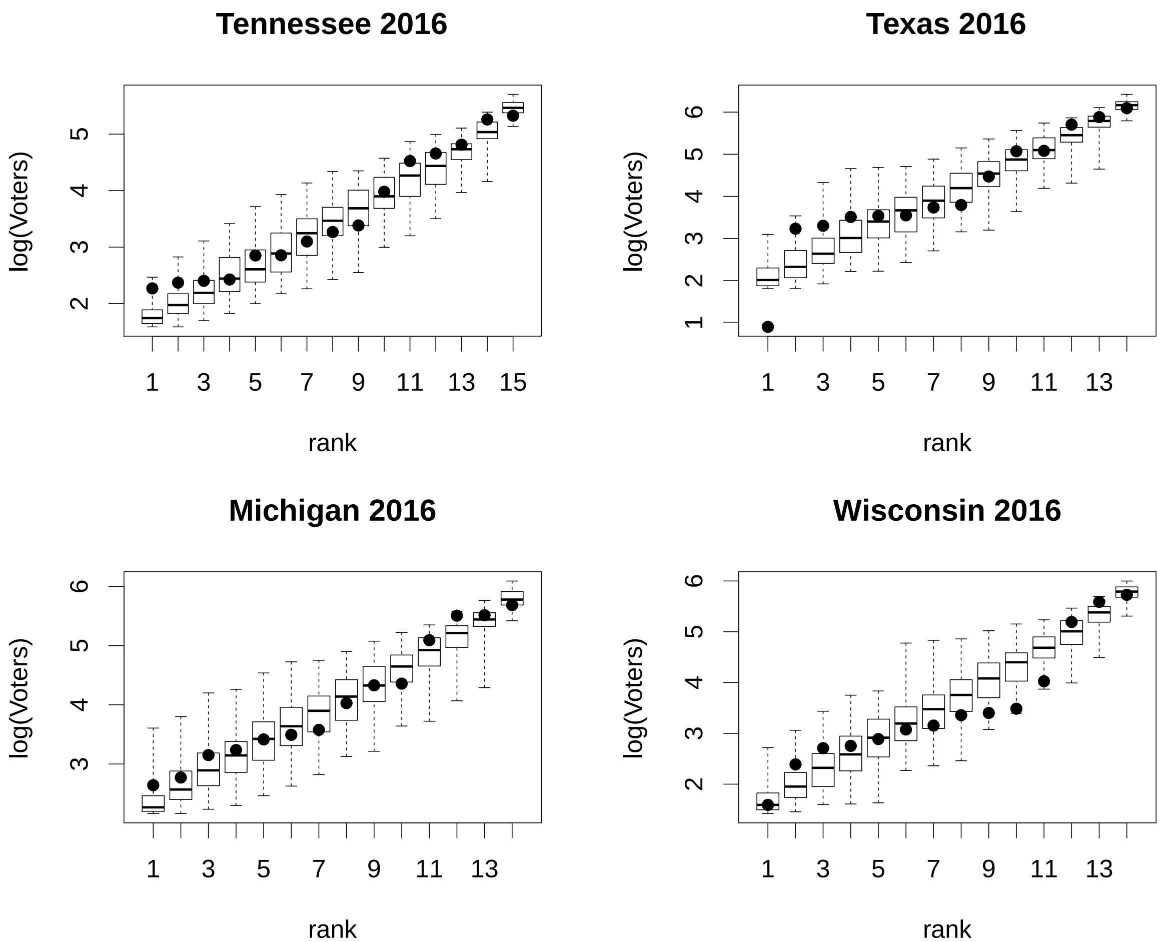


Figure 60: Election US (republicans), 2016. Note that, in case of Texas, z was estimated according to an outlier (boxplot of 100 realizations og the model, bullets: data).