



Logit Scaling: A General Method for Alignment in Microsimulation models

Peter Stephensen

Danish research institute of applied economic modelling, DREAM
Copenhagen, Denmark
psp@dreammodel.dk

ABSTRACT: A general and easily implemented method for alignment in microsimulation models is proposed. Most existing alignment methods address the binary case and they typically put special emphasis on one of the alternatives rather than treating all alternatives in a symmetric way. In this paper we propose a general mathematical foundation of multinomial alignment, which minimizes the relative entropy in the process of aligning probabilities to given targets. The method is called Logit Scaling. The analytical solution to the alignment problem is characterized and applied in deriving various properties for the method. It is demonstrated that there exists an algorithm called Bi-Proportional Scaling that converges to the solution of the problem. This is tested against two versions of the Newton-Raphson-algorithm, and it is demonstrated that it is at least twice as fast as these methods. Finally, the method is not just computational efficient but also easy to implement.

KEYWORDS: Microsimulation, Alignment, Bi-proportional scaling, Relative entropy, Kullbeck-Leibler Information Criterion, Newton-Raphson, Logit scaling

JEL classification: C1, C5, C81

I INTRODUCTION

Alignment is the core facility for controlling microsimulation models. Microsimulation models are large, elaborate models that need such a tool at the macro level. Alignment works by manipulating the probabilities of the model, aiming to do *model calibration*, *comparative analysis* (static or dynamic) and/or to *eliminate Monte-Carlo noise*.

In a more technical sense, alignment can be said to do one of two things: *mean-correction* or *variance-elimination*¹. The focus of this paper is mean-correction, meaning the process by which you manipulate the probabilities of the individuals such that i) you hit a predefined mean/average propensity for the entire group, and ii) the manipulation is as gentle as possible (discussed below). Mean-correction alignment makes it possible to calibrate the model and do comparative analysis. The other kind of alignment, here called variance-elimination, is equivalent to the above mentioned elimination of Monte-Carlo noise. In microsimulation models, events are typically modeled by drawing a uniformly distributed number and comparing it to a probability associated with the event occurring. This Monte-Carlo technique introduces noise into the model. Such noise can be eliminated from the model, using alignment techniques (Bækgaard, 2002).

Most existing alignment methods address the binary case and they typically put special emphasis on one of the alternatives rather than treating all alternatives in a symmetric way (for a review, see Li & O'Donoghue, 2014). In this paper we propose a more general mathematical foundation of multinomial alignment. Our mean-correction alignment method is called *Logit Scaling*, and is defined as a method that minimizes the *information loss* in the adjustment process. The analytical solution to the alignment problem is characterized and furthermore applied in deriving various properties for the method. It is demonstrated that there exists an algorithm called Bi-Proportional Scaling that converge to the solution of the problem. This is tested against two versions of the Newton-Raphson-algorithm, and it is demonstrated that it is at least twice as fast as these methods. Finally, it is argued, that the method is very easy to implement.

The practical implementation of Logit Scaling is very simple. Assume N individuals can be in one of A alternative states, and that you have probabilities p_{ia}^0 for all individuals $i \in \{1, \dots, N\}$ and alternatives $a \in \{1, \dots, A\}$. We necessarily have that $\sum_a p_{ia}^0 = 1$ for all i and $p_{ia}^0 \geq 0$ for all i and a . In the alignment process you would like to find new probabilities p_{ia} such that the aggregated number of individuals in the various states satisfies certain targets:

$$\sum_i p_{ia} = \bar{X}_a, 1 \leq a \leq A.$$

The algorithm we use is called Bi-Proportionate Scaling and is demonstrated in Algorithm 1. Initially let $p_{ia} = p_{ia}^0$. Think about p_{ia} as a matrix with N rows and A columns. Scale all columns such that the alignment targets are satisfied ($\sum_i p_{ia} = \bar{X}_a$). Then scale all rows such that $\sum_a p_{ia}^0 = 1$ for all i . Repeat this until the system converges. In real-world problems the algorithm converges very quickly even for many individuals and states.

Algorithm 1 The Bi-Proportionate Scaling algorithm used in Logit Scaling.

Require: Initially $p_{ia} = p_{ia}^0$

Ensure: $\sum_i p_{ia} = \bar{X}_a$

- 1: **while** not converged **do**
 - 2: Scale columns with γ_a 's: $p_{ia} := \gamma_a p_{ia}$ such that $\sum_i p_{ia} = \bar{X}_a$
 - 3: Scale rows with α_i 's: $p_{ia} := \alpha_i p_{ia}$ such that $\sum_a p_{ia} = 1$
 - 4: **end while**
-

2 WHAT IS A GOOD ALIGNMENT METHOD?

We would like to start by defining what a good mean-correction alignment method is. The points in the following list ensures that the alignment method is theoretically sound and useable at the same time:

1. Adjust correctly to the target means
2. Retain the original shape of the probability distribution
3. Retain zero probabilities
4. Symmetric formulation
5. Multinomial alignment
6. Ability to work well in a logit environment
7. Computer efficiency
8. Easy implementation

Points 1) and 2) are similar to the first two points in Li & O'Donoghue (2014). The idea of mean-correction alignment is to be able to change the mean of the probability distribution. Therefore the method should be precise when it comes to adjusting to a new target (point 1). Let's try to make this more precise. Consider a population of N individuals, and a number A of alternatives the individuals can be in. Let p_{ia} be individual i 's probability of entering alternative a ($i = 1, \dots, N$ and $a = 1, \dots, A$). In a binary model $A = 2$. Because we are working with probabilities we should necessarily have that

$$\sum_a p_{ia} = 1, i = 1, \dots, N$$

Let X_a be the number of persons in alternative a (such that $\sum_a X_a = N$). The expected number of persons

in alternative a is given by:

$$E[X_a] = \sum_i p_{ia}, \quad a = 1, \dots, A$$

We can now define mean-correction alignment:

Definition 1. Let p_{ia}^0 ($i = 1, \dots, N$ and $a = 1, \dots, A$) be the initial probabilities in a population. A *mean-correction alignment* to a target $(\bar{X}_1, \dots, \bar{X}_A)$ is a set of aligned probabilities p_{ia} , ($i = 1, \dots, N$ and $a = 1, \dots, A$) such that:

$$\sum_i p_{ia} = \bar{X}_a, \quad a = 1, \dots, A \quad (2.1)$$

$$\sum_a p_{ia} = 1, \quad i = 1, \dots, N \quad (2.2)$$

$$p_{ia} \geq 0, \quad i = 1, \dots, N, \quad a = 1, \dots, A \quad (2.3)$$

Observe that this definition only involves the aligned probabilities. Nothing is said about the link between the initial and the aligned probabilities. This is the subject of point 2: a very basic demand for an alignment method is that it makes as little harm as possible to the original probabilities, in the sense that we strive to retain the shape of the original distribution. A way to handle this is to define a distance measure that measures the difference between the initial probabilities and the aligned probabilities. Let P be a $N \times A$ matrix containing the p_{ia} -elements. Let $D(P, P^0)$ be our measure. A precise version of point 2 would be to minimize $D(P, P^0)$ given the restrictions in Definition 1. We will return to this optimization problem later.

The third point in our list deals with zero probabilities. Saying that a probability is zero is a very strong statement. It means that something is impossible. An alignment method should therefore retain zeroes such that what is impossible before the alignment is also impossible after the alignment.

The fourth point on the list is *symmetric formulation*. A good example of an asymmetric alignment method is the often used *multiplicative scaling* (Li & O'Donoghue, 2014). According to this kind of alignment you choose one of the alternatives (let's say alternative 1) and scale the probabilities to the desired level:

$$p_{i1} = \lambda p_{i1}^0$$

where the scaling factor λ is defined by

$$\sum_i p_{i1} = \sum_i \lambda p_{i1}^0 \equiv \bar{X}_1$$

If we have two alternatives, the probability of the other alternative is given residually

$$p_{i2} = 1 - p_{i1}$$

This is clearly asymmetric. If we had chosen to scale the probabilities of the second alternative instead, we would have had another alignment. This is not satisfying as the alignment method introduces arbitrariness into the analysis. This asymmetry seems not to be a well-known property in the microsimulation community. As an

example it is not mentioned in Li & O'Donoghue (2014). We give an graphical example in section 3.2.

The only way to make sure that the alignment formulation is symmetric, is to make a definition based on all alternatives. A good alignment method should therefore depend on *all* the initial probabilities and some parameter vector θ_a

$$p_{ia} = f_a(p_{i1}^0, \dots, p_{iA}^0; \theta_a), a = 1, \dots, A$$

This is typically not the case in the most used alignment methods (Li & O'Donoghue, 2014).

A good alignment technique should generalize easily to the *multinomial* case ($A > 2$). As mentioned above, most alignment methods are binary. Multinomial problems are often solved by a sequence of binary alignments. This is also a source of arbitrariness. If we change the order of the binary alignments we potentially get a different total alignment. It is therefore preferable to have a method that is designed to the general multinomial case.

The final 3 points on the list have to do with the usability of the method. The logit specification is one of the most used methods to estimate probabilities. It is therefore a nice feature if an alignment method works well with this specification². Finally it is important for the usability of a method to have both computer efficiency and ease of implementation. These two goals can be conflicting goals. As will be demonstrated later, this is not the case here.

3 A MULTINOMIAL ALIGNMENT METHOD

You can think about alignment as a method for manipulating probability distributions. Define

$$q_{ia} \equiv \frac{1}{N} p_{ia}$$

where N is the number of individuals. $Q = \{q_{ia} | i = 1, \dots, N; a = 1, \dots, A\}$ is a joined discrete probability distribution simultaneously describing preferences (the A alternatives) and heterogeneity (the N individuals)³. Let Q^0 be the initial probability distribution. According to definition 1, a mean-correction alignment to a target $(\bar{X}_1, \dots, \bar{X}_A)$ is a new probability distribution Q satisfying:

$$\sum_i q_{ia} = \frac{1}{N} \bar{X}_a, a = 1, \dots, A$$

$$\sum_a q_{ia} = \frac{1}{N}, i = 1, \dots, N$$

$$q_{ia} \geq 0, i = 1, \dots, N, a = 1, \dots, A$$

We would like to satisfy these linear restrictions with as little change in the distribution as possible.

Our way of solving this problem is inspired by methods developed in finance during the mid 90s (Stutzer, 1996; Kitamura & Stutzer, 1997; Buchen & Kelly, 1996). The basic idea is to minimize the relative entropy between the two distributions, subject to the additional restrictions imposed on the new distribution. In Stutzer (1996)

the goal is to modify a nonparametric predictive distribution for the price of an asset to satisfy the martingale condition associated with risk-neutral pricing. In Kitamura & Stutzer (1997) the goal is to provide an alternative to generalized method of moments estimation in which the moment conditions holds exactly. In Buchen & Kelly (1996) it is demonstrated how to estimate a distribution of an underlying asset from a set of option prices. The method has been used by Robertson et al. (2005) for VAR model forecasting. For an overview of the use of entropy in finance, see Zhou et al. (2013).

The entropy measure used in the above papers is the Kullback-Leibler Information Criterion (KLIC) distance from Q_0 to Q (White, 1982):

$$\begin{aligned} D(Q, Q^0) &= \sum_i \sum_a q_{ia} \log \left(\frac{q_{ia}}{q_{ia}^0} \right) \\ &= \frac{1}{N} \sum_i \sum_a p_{ia} \log \left(\frac{p_{ia}}{p_{ia}^0} \right) \end{aligned} \quad (3.1)$$

This is also called the *Relative Entropy*. It is well-known that $D(Q, Q^0) \geq 0$, with equality if and only if $Q = Q^0$. The KLIC measures the information gained when one revises one's beliefs from the prior probability distribution Q^0 to the posterior probability distribution Q . In our use of the concept we therefore wish to minimize the *information loss* from moving away from the original distribution. The KLIC is a special case of a broader class of divergences called f-divergences as well as the class of Bregman divergences. It is the only such divergence over probabilities that is a member of both classes.

The KLIC is not a true *metric* as it is not symmetric ($D(Q, Q^0) \neq D(Q^0, Q)$). Even so, it is a *premetric* and generates a topology on the space of probability distributions. To see that it actually works as it should, let us look at the binary case. Define:

$$D_i = \sum_a q_{ia} \log \left(\frac{q_{ia}}{q_{ia}^0} \right)$$

In the binary case we have

$$D_i = q_{i1} \log \left(\frac{q_{i1}}{q_{i1}^0} \right) + (1 - q_{i1}) \log \left(\frac{1 - q_{i1}}{1 - q_{i1}^0} \right)$$

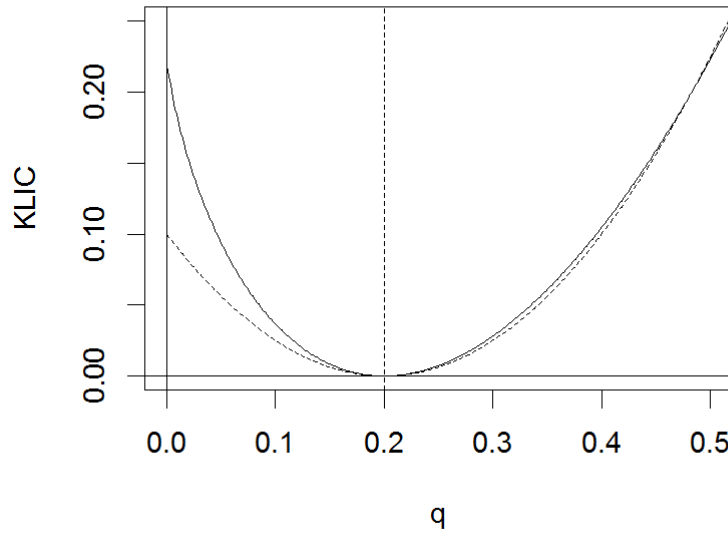
For which value of q_{i1} is this function minimized? For D_i to be a distance measure we would prefer $q_{i1} = q_{i1}^0$ to minimize the function, and this is surely the case. Differentiation yields:

$$\frac{\partial D_i}{\partial q_{i1}} = \log \left(\frac{q_{i1}}{q_{i1}^0} \right) - \log \left(\frac{1 - q_{i1}}{1 - q_{i1}^0} \right)$$

and $\frac{\partial D_i}{\partial q_{i1}} = 0$ if and only if $q_{i1} = q_{i1}^0$. In Figure 3.1 an example is shown. It is assumed that $q_{i1}^0 = 0.2$. For comparison the dashed curve shows $a (q_{i1} - q_{i1}^0)^2$ where a is chosen such that the curves are close to the right of q_{i1}^0 . It can be seen that KLIC acts much like quadratic distance to the right of the minimum, but has more curvature to the left.

We would like to minimize the information loss D given the restrictions (2.1), (2.2) and (2.3) from definition

Figure 3.1: Example of binary KLIC (solid line) compared to quadratic distance. It is assumed that $q_{i1}^0 = 0.2$.



1. The restriction (2.1) makes sure that we satisfy 1) in our list (ability to adjust correctly to the target means). The fact that we minimize the information loss satisfies 2) in the list (ability to retain the original shape of the probability distribution). 'Symmetric formulation' and 'multinomial alignment' follows automatically (point 4 and 5).

We can now derive our new alignment method:

Definition 2. Let p_{ia}^0 ($i = 1, \dots, N$ and $a = 1, \dots, A$) be the initial probabilities in a population. A *Relative-Entropy-Minimizing Mean-Correction Alignment* to a target $(\bar{X}_1, \dots, \bar{X}_A)$ is a set of aligned probabilities p_{ia} that minimize (3.1), given the restrictions (2.1), (2.2) and (2.3).

And we can prove the theorem:

Theorem 3. Let p_{ia}^0 ($i = 1, \dots, N$ and $a = 1, \dots, A$) be the initial probabilities in a population. In a *Relative-Entropy-Minimizing Mean-Correction Alignment* the aligned probabilities p_{ia} are given by:

$$p_{ia} = \frac{e^{\varphi_a} p_{ia}^0}{\sum_s e^{\varphi_s} p_{is}^0}$$

where the parameters φ_a satisfy:

$$\sum_a \varphi_a = 0$$

Proof. (White, 1982) Define the Lagrange-function L :

$$L = \sum_i \sum_a p_{ia} (\log(p_{ia}) - \log(p_{ia}^0)) - \sum_i \lambda_i \left(\sum_a p_{ia} - 1 \right) - \sum_a \varphi_a \left(\sum_i p_{ia} - \bar{X}_a \right)$$

The first order derivatives are given by

$$\frac{\partial L}{\partial p_{ia}} = (\log(p_{ia}) - \log(p_{ia}^0)) + 1 - \lambda_i - \varphi_a = 0$$

such that:

$$p_{ia} = e^{\lambda_i - 1} e^{\varphi_a} p_{ia}^0 \quad (3.2)$$

Sum over the alternatives a (using (2.2)):

$$\sum_a p_{ia} = \sum_a e^{\lambda_i - 1} e^{\varphi_a} p_{ia}^0 = 1$$

such that

$$e^{\lambda_i - 1} = \left[\sum_a e^{\varphi_a} p_{ia}^0 \right]^{-1}$$

Inserting this into (3.2) yields:

$$p_{ia} = \frac{e^{\varphi_a} p_{ia}^0}{\sum_s e^{\varphi_s} p_{is}^0}$$

Observe that the Lagrange-parameters φ_a are not unique, as

$$\frac{e^{\varphi_a} p_{ia}^0}{\sum_s e^{\varphi_s} p_{is}^0} = \frac{e^{x+\varphi_a} p_{ia}^0}{\sum_s e^{x+\varphi_s} p_{is}^0}$$

We can therefore assume that

$$\sum_a \varphi_a = 0$$

□

Observe that $p_{ia} \in [0, 1]$ and that 3) is satisfied (retaining zero).

3.1 LOGIT SCALING

We will actually prefer to give our method the catchier name *Logit Scaling*. The method fits perfectly a situation where the initial probabilities are supplied from a logit-model (point 6 in the list). Assume the probabilities are given by

$$p_{ia}^0 = \frac{e^{V_{ia}^0}}{\sum_s e^{V_{is}^0}}$$

where V_{ia}^0 are some logit-utilities. Then according to theorem 3:

$$p_{ia} = \frac{e^{\varphi_a + V_{ia}^0}}{\sum_s e^{\varphi_s + V_{is}^0}}$$

The alignment is done by adding alternative-specific constants to the logit-utilities. As a corollary to this, if the preferences of the agents change in an additive way, the probabilities can be corrected with alignment. This is similar to the result in Li & O'Donoghue (2014) where it is shown that a “biased intercept” in a logit model can be corrected with some of the alignment methods (“Sidewalk hybrid with non linear adjustment” and “Sort by the difference between logistic adjusted predicted probability and random number”). Here we give a mathematical explanation.

After an alignment, the scaling parameters φ_a can be calculated and used to solve the following problem: assume we have a logit model describing some events in our microsimulation model. In the base run of the model we align these logit probabilities to given macro levels. In an alternative analysis we would like to calculate the behavioral responses using the logit model. But how can we do that when we have manipulated our logit probabilities in the alignment process?

Assume the logit model is given by

$$p_{ia}^0 = \frac{e^{\beta_a x_i^0}}{\sum_s e^{\beta_s x_i^0}}$$

where x_i^0 is a vector characterizing individual i and β_a is a vector of logit-parameters for each alternative a . Let the aligned probabilities be:

$$p_{ia} = \frac{e^{\varphi_a + \beta_a x_i^0}}{\sum_s e^{\varphi_s + \beta_s x_i^0}} \quad (3.3)$$

Now assume we would like to analyze the effect of changing x_i^0 to x_i for all individuals i . Having calculated the φ'_a s we can calculate the new probabilities \hat{p}_{ia} :

$$\hat{p}_{ia} = \frac{e^{\varphi_a + \beta_a x_i}}{\sum_s e^{\varphi_s + \beta_s x_i}}$$

These probabilities will give the best possible representation of the marginal effects of the changed characteristics of the individuals.

If the aligned probabilities p_{ia} are known, it is possible to calculate the φ'_a s. It can be derived that, for any i :

$$\varphi_a = \log \left(\frac{p_{ia}}{p_{ia}^0} \right) - \frac{1}{A} \sum_s \log \left(\frac{p_{is}}{p_{is}^0} \right) \quad (3.4)$$

This is so-called *centered log-ratios* known from compositional data analysis (Aitchison, 1986). Even if we are not doing the alignment via a direct calculation of the φ'_a s (an idea we will introduce in the next section) we can calculate them from (3.4).

In the binary case ($A = 2$) it can be proven that

$$\varphi_a = \frac{1}{2} \left(\log \left(\frac{p_{ia}}{1 - p_{ia}} \right) - \log \left(\frac{p_{ia}^0}{1 - p_{ia}^0} \right) \right)$$

such that the scaling factors are calculated from changes in log-odd-ratios.

3.2 An example

Let us consider a simple example. We have two alternatives ($A = 2$; death or survive) and two individuals ($N = 2$). We assume that the death probability of the first person is $p_1 = 0.2$ (such that the probability of survival is 0.8) and the death probability of the second person is $p_2 = 0.4$. The expected number of deaths in this tiny microsimulation model is $0.2 + 0.4 = 0.6$. Let us assume we want to increase the expected number of deaths to 0.85. We will do this by alignment.

The first method we will use is the popular *Multiplicative Scaling* (Li & O'Donoghue, 2014). The initial point is A in the figure. The alignment process should move us to some point on the downward sloping curve defined by $p_1 + p_2 = 0.85$. With multiplicative scaling we have two possibilities⁴: scaling up the probabilities of death or scaling down the probabilities of survival. If we scale up the probabilities of death we move from A to B. If we scale down the probabilities of survival we move from A to C. As we can see there is a significant difference in the resulting alignments.

Another problem associated with multiplicative scaling is that the aligned probabilities can exceed 1. If the death probability is scaled up to a certain level we reach the point Z. Here the probability of death for person 2 is 1. Any further scaling will imply a death probability larger than 1. Likewise, if the probabilities of survival is scaled enough, we reach the point Y.

Now let us consider *Logit Scaling*. In the binary case we have from theorem 3 that:

$$p_i = \frac{e^\varphi p_i^0}{e^\varphi p_i^0 + e^{-\varphi} (1 - p_i^0)} \quad (3.5)$$

The microsimulation model DYNACAN uses an algorithm called 'Sidewalk with Non-Linear Transformation' (Li & O'Donoghue, 2014). This non-linear transformation is given by:

$$p = \frac{ap^0}{1 + (a - 1)p^0}, \quad a > 0.$$

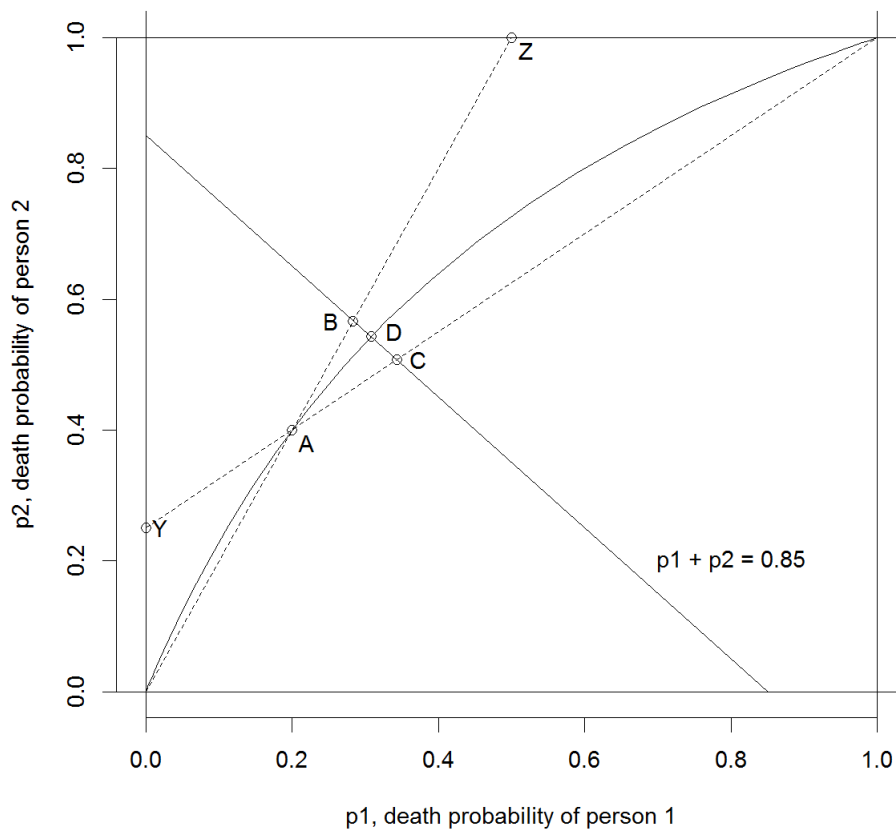
We can re-write (3.5) to:

$$p_i = \frac{e^{2\varphi} p_i^0}{1 + (e^{2\varphi} - 1)p_i^0}$$

such that the transformations are identical for

$$a = e^{2\varphi}$$

Figure 3.2: Alignment methods



In the figure the solid non-linear curve shows all possible alignments for varying values of φ . The current alignment is the point D. Observe that compared to multiplicative scaling, the method is a compromise between point A and B. Also observe that the system can be aligned to any number of deaths between 0 and $N = 2$.

4 BI-PROPORTIONAL SCALING: AN ALGORITHMIC SOLUTION

So far the alignment method seems to have nice theoretical properties, but how about the practical side of the equation? If we substitute the results of theorem 3 into equation (2.1), we get the following problem: choose parameters φ_a , $a = 1, \dots, A$, such that:

$$\sum_a \varphi_a = 0$$

$$\sum_i \frac{e^{\varphi_a} p_{ia}^0}{\sum_s e^{\varphi_s} p_{is}^0} = \bar{X}_a, a = 1, \dots, A - 1$$

This is a system of A non-linear equations with A unknowns. As A is typically small (2 to 10), this is not a big equation system. But from a computational point of view it is a problem that we sum over the number of individuals i . This can be a very big number (up to millions). We therefore need a fast algorithm.

	Time (Sec.)	φ_1	φ_2	φ_3
Newton-Raphson (Standard)	3.94	0.53841805	-0.58964392	0.00557946
Newton-Raphson (Optimized)	1.45	0.54167153	-0.59175494	0.00659708
Bi-Proportional Scaling	0.60	0.53841807	-0.58964390	0.00557951

Table 1: Comparing Bi-proportional scaling with two Newton-Raphson algorithms ($N = 1.000.000$, $A = 4$). The centered log-ratios φ'_o s are the output of the Newton-Raphson algorithms. In the case of Bi-Proportional Scaling they are calculated using (3.4).

The standard computational way to solve a set of non-linear equations is to use the Newton-Raphson algorithm. This algorithm works by making a local linear approximation of the system (the Jacobian matrix). This linear system is solved (the Jacobian matrix is inverted), and by repeating the process it is possible to converge to the solution of the non-linear set of equations.

Using the Newton-Raphson algorithm implies the calculation of the $A \times A$ Jacobian matrix in each iteration, where each element in this matrix is a sum over all individuals. A considerably faster approach is *Bi-proportional scaling*. Look at the original problem: minimize (3.1) given (2.1) and (2.2). This is actually a *Matrix Balancing Problem* (Schneider & Zenios, 1990). Instead of the initial matrix P^0 we should choose another matrix P , such that the difference between the matrices is as small as possible, and such that the row sums are given (see 2.2) and the column sums are given (see 2.1). It is well known (McDougall, 1999) that if the measure of difference between the matrices is Relative Entropy, then there exists a simple algorithm that converges to the solution of the matrix balancing problem. This algorithm is called bi-proportional scaling and is very simple and fast. You start out with the original matrix P^0 . In this matrix the row sums are 1 because we work with probabilities. The column sums are equal to the expected numbers of outcomes distributed on the A alternatives. First you scale the columns such that the column sums are correct according to the restrictions, i.e. the expected numbers are equal to the alignment targets. Then you scale the rows such that the column sums are correct (sums to 1 again). But then the column sums are not correct any more. If you repeat this process again and again, you will realize that the errors in the column sums will decrease dramatically, and the system will converge. In real world problems the convergence is very fast (often less than 10 iterations). We therefore have a very surprising result: our very non-linear problem can be solved by simple iterative scaling.

The method is tested against two versions of the Newton-Raphson algorithm, a standard version and an optimized version (see table 1). The reason for this comparison is twofold: first, it is proven that Bi-proportional scaling is faster than the standard method (even when it is optimized), and second, it is demonstrated that the method yields the same results.

In the standard version of the Newton-Raphson algorithm, the Jacobian matrix is calculated based on abstract differentiation of the non-linear equation system. In each iteration the elements in the matrix is calculated by an average over all individuals. The matrix is inverted using the C#-library Math.Net. In the optimized version, the procedure is the same except for the calculation of the Jacobian matrix. The elements in the matrix is averaged over a sample of 10 pct. of the individuals. This improves the speed, only slightly compromising precision.

The artificial population is 1.000.000 individuals with 4 alternatives. The artificial set of probabilities are calculated by drawing 1 mill. numbers from each of 4 normal distributions⁵. If $(x_{i1}, x_{i2}, x_{i3}, x_{i4})$ are these 4

number for individual i , the probabilities are calculated as:

$$p_{ia} = \frac{x_{ia}}{\sum_{s=1}^4 x_{is}}$$

The result of the test is given in Table 1. Bi-proportional scaling ran twice as fast as the optimized Newton-Raphson, and more than 6 times faster than the standard Newton-Raphson. If we look at the centered log-ratios we see that these are practically identical for the standard Newton-Raphson and Bi-Proportional Scaling. As mentioned the speed of the optimized Newton-Raphson resulted in a somewhat lower precision. This precision cost is not observed for Bi-Proportional Scaling.

The final point on the list is ease of implementation. It took more than a week to develop and test the Newton-Raphson algorithms. When the idea came to the author, it took half an hour to code bi-proportional scaling, showing the power of the approach.

5 CONCLUSION

A general mathematical foundation of multinomial alignment in microsimulation models is derived. The alignment problem is characterized as the solution to an optimization problem. The analytical solution to this problem is characterized and applied in deriving various properties for the method. It is demonstrated that there exists an algorithm called *Bi-Proportional Scaling* that converges to the solution of the optimization problem, that the new method is more general than existing methods, and that it is fast and easy to implement.

ACKNOWLEDGMENT

The author would like to thank Joao Miguel Ejarque, Michael Andersen, Marianne Frank Hansen and two anonymous referees for many thoughtful comments and edits. We would also like to thank the The Knowledge Centre for Housing Economics under Realdania for generous financial support.

REFERENCES

- Aitchison, J. (1986). *The statistical analysis of compositional data*. London: Chapman and Hall London.
- Bækgaard, H. (2002). Micro-macro linkage and the alignment of transition processes. *Technical paper*(25).
- Buchen, P. W., & Kelly, M. (1996). The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis*, 31(01), 143–159.
- Kitamura, Y., & Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments estimation. *Econometrica: Journal of the Econometric Society*, 861–874.

- Li, J., & O'Donoghue, C. (2014). Evaluating binary alignment methods in microsimulation models. *Journal of Artificial Societies and Social Simulation*, 17(1), 15.
- McDougall, R. A. (1999). Entropy theory and RAS are friends. *GTAP Working Papers*, 6.
- Robertson, J. C., Tallman, E. W., & Whiteman, C. H. (2005). Forecasting using relative entropy. *Journal of Money, Credit, and Banking*, 37(3), 383-401.
- Schneider, M. H., & Zenios, S. A. (1990). A comparative study of algorithms for matrix balancing. *Operations research*, 38(3), 439-455.
- Stutzer, M. (1996). A simple nonparametric approach to derivative security valuation. *The Journal of Finance*, 51(5), 1633-1652.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, 1-25.
- Zhou, R., Cai, R., & Tong, G. (2013). Applications of entropy in finance: A review. *Entropy*, 15(11), 4909-4931.

¹Or a more general term for both: *moment correction*, as it corrects the first two moments of the probability distribution.

²Admittedly this point is added because our method works well with logits.

³ Q is actually a probability distribution as $0 \leq q_{ia} \leq 1$ and $\sum_i \sum_a q_{ia} = 1$.

⁴This asymmetry seems not to be a well-known in the microsimulation community. As an example it is not mentioned in Li & O'Donoghue (2014)

⁵The distributions are: $N(-3.0, 0.8)$, $N(-1.0, 0.5)$, $N(0, 0.5)$ and $N(-0.2, 0.8)$ where the first number is mean and the second is variance.