

RESEARCH ARTICLE

# The Efficiency of Split Panel Designs in an Analysis of Variance Model

Xin Liu<sup>1</sup>, Wei-Guo Wang<sup>2\*</sup>, Hai-Jun Liu<sup>3</sup>

**1** School of Economics and Management, Liaoning Shihua University, Fushun, China, **2** School of Economics, Dongbei University of Finance and Economics, Dalian, China, **3** School of Marxism, Dalian Maritime University, Dalian, China

\* [wwgstudent@126.com](mailto:wwgstudent@126.com)



OPEN ACCESS

**Citation:** Liu X, Wang W-G, Liu H-J (2016) The Efficiency of Split Panel Designs in an Analysis of Variance Model. PLoS ONE 11(5): e0154913. doi:10.1371/journal.pone.0154913

**Editor:** Boris Podobnik, University of Rijeka, CROATIA

**Received:** November 21, 2014

**Accepted:** April 21, 2016

**Published:** May 10, 2016

**Copyright:** © 2016 Liu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the References section of the manuscript or are owned by Informart, a market research agency. Interested researchers may contact <http://www.sciencedirect.com/science/article/pii/S030440769090013J> for access to data from Informart.

**Funding:** This work is supported financially by talent scientific research fund of LSHU, Social Science Foundation of China (No. 15BKS073), Natural Science Foundation of China (No. 71171035, 71271045, 71571033, 11301060), the Program for Discipline Construction in DUFF (No. XKT-201411) and Outstanding scientific innovation talents program of DUFE (no.DUFE2014R20). The funders had no

## Abstract

We consider split panel design efficiency in analysis of variance models, that is, the determination of the cross-sections series optimal proportion in all samples, to minimize parametric best linear unbiased estimators of linear combination variances. An orthogonal matrix is constructed to obtain manageable expression of variances. On this basis, we derive a theorem for analyzing split panel design efficiency irrespective of interest and budget parameters. Additionally, relative estimator efficiency based on the split panel to an estimator based on a pure panel or a pure cross-section is present. The analysis shows that the gains from split panel can be quite substantial. We further consider the efficiency of split panel design, given a budget, and transform it to a constrained nonlinear integer programming. Specifically, an efficient algorithm is designed to solve the constrained nonlinear integer programming. Moreover, we combine one at time designs and factorial designs to illustrate the algorithm's efficiency with an empirical example concerning monthly consumer expenditure on food in 1985, in the Netherlands, and the efficient ranges of the algorithm parameters are given to ensure a good solution.

## Introduction

Split panel combines advantages of other sampling methods (including repeated, cross-sectional, and rotating sample) and provides us with rich, convenient, and practical information, being widely applied in many fields [1]. In experiments with economic principles survey, researchers typically consider statistical models that allow complex relationships. Due to the complex statistical model requirements, data should be collected to estimate the statistical model parameters. Being a new sample type that combines the advantages of the other three basic samples, split panel is used to provide rich data for complex statistical model. Since wide application of micro-economic data, panel conditioning, and panel nonresponse become more important in econometrics, as well documented in literature [1], split panel, as a combination of a panel and a repeated or rotating panel, uses changing samples to recruit from replacements for panel conditioning and panel nonresponse [2–3]. In most fields, such as finance, labor economics, and political economy, the collection of data is characterized by high costs. Split panel

role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

has the advantages of the flexibility of the cross-section change sizes and continually updating information [4–5]. Therefore, it is very important to use split panel and the optimal sample design to obtain as much information as possible from a given budget. In recent years, the theory of split panel sample has witnessed theoretical advances and applications across disciplines of pure and applied sciences, and it will be widely used in the future [6–7]. However, limited attention has been paid to the analysis of split panel design efficiency recently. In the early literature, the estimation of a time-dependent mean from several kinds of rotating samples, that is, the special form of split panel and the resulting variances have been examined by Patterson [8] and Eckler [9]. It has been documented that the optimal design of the sample depends on the parameter of interest (see [10], pp. 152). On this basis, Nijman et al. [11] determined the optimal split panel design, that is, how to choose the optimal proportion of a given budget that can be spent on the collection of a series of cross-sections to minimize split panel design efficiency. However, in sampling, we need to obtain the optimal proportion of a series of cross-sections in all samples, and it cannot be obtained accurately by the proportion of the budget that can be spent on the collection of a series of cross-sections [11]. Consequently, we cannot save sampling costs according to [11]. On the other hand, the split panel design optimization algorithm is not given in [11]. For researchers and practitioners to solve for the optimal split panel design, they need to select or design the appropriate optimization algorithm and to calculate the optimal proportion by the optimization theory. In sampling, this will decrease the efficiency of calculating the optimal proportion and reduce the accuracy of the solution. Hence, it is not attractive to design split panel in the research framework of [11].

In this paper, the goal is to minimize the efficiency of split panel design in the analysis-of-variance model, when one needs to determine the optimal proportion of a series of cross-sections in all samples, when the optimal proportion of a series of cross-sections can be applied directly in sampling. This is an extension of [11], and the main contributions can be summarized as follows. First, we show how to choose the proportion of a series of cross-sections, in all samples, to minimize the variances of estimators in the analysis-of-variance model, irrespective of the parameters of interest and budget. In particular, we present the relative efficiency of estimators, based on the split panel to an estimator based on a pure panel or a pure cross-section. Second, we transform the efficiency of split panel design under a budget constraint for the nonlinear integer optimization (difficult to solve by mature optimal algorithms). The simulated annealing algorithm has the advantages of guaranteeing global optimization, selecting the initial solution randomly, and being simple and practical [12]. However, when the simulated annealing algorithm is used to solve the constrained nonlinear integer optimization associated with the efficiency of split panel design under a budget constraint, it is difficult to combine parameters, such as the inner iteration number, the initial temperature, and the temperature decrease rate, in order to get the best performance of the algorithm. Hitherto, there is no theoretical method to solve this problem. Therefore, in this paper we design an efficient algorithm, based on the simulated annealing algorithm, to solve the constrained nonlinear integer optimization of the split panel design efficiency under budget constraint. In the context of numerical modeling, sensitivity analysis studies how different values of an independent variable impact a particular dependent variable, under a given set of assumptions. It has been widely applied to many fields such economics, engineering, ecology, etc. The modelers can determine, by sensitivity analysis, whether the parameters of the model or algorithm give reliable predictions. Hence, third, we introduce sensitivity analysis to appraise the parameters of the proposed algorithm. The one-at-a-time design (OATD) method is one of the most common approaches for the effect on the output [13–16] and it is frequently used as the modeler immediately knows which input factor is responsible for the failure, in case of model failure [17]. Yet, the OATD method cannot be used if two factors are interdependent, because it only studies the effect of

one [variable](#) at a time. The factorial design (FD) method, which is used to study the effects several factors have on a response, and the interactions between the factors for varying levels of all factors at the same time, is different from the OATD method. As such, the OATD and FD methods are chosen to analyze the effect of parameters and compensate the deficiency of a single method [16]. On the other hand, the simulated annealing algorithm has no special requirement and its performance cannot be changed with different examples [12]. Hence, with an empirical example concerning monthly consumer expenditure on food in 1985, in the Netherlands, we combine the OATD and FD methods to analyze the algorithm designed in this paper. The result are the efficient ranges of the algorithm parameters are a good solution (i.e., the accurate optimal proportion). Therefore, the research results in this paper would be useful to both researchers and practitioners in sampling.

This paper is organized as follows: section 2, based on the analysis-of-variance model, transforms the efficiency of split panel design into a nonlinear optimization; in section 3, the efficiency of split panel design, irrespective of interest and budget, is discussed; in section 4, we consider the efficiency of split panel design given a budget constraint and design an efficient algorithm based on simulated annealing to solve the resulting constrained nonlinear integer optimization; section 5 combines the OATD and FD methods to illustrate the algorithm’s efficiency, with an empirical example of food monthly consumer expenditure in 1985, in the Netherlands, and the efficient ranges of the algorithm parameters are given to ensure a good solution; and section 6 concludes the paper.

## Materials and Methods

### Theoretical results of parameter estimators variances

In this paper, we consider the split panel design efficiency by minimizing the best linear estimator variance of the linear combinations  $\sum_{t=1}^T \phi_t \hat{\beta}_t$  of the period means  $\hat{\beta}_t$  in the analysis of the variance model:

$$y_{it} = \beta_t + \alpha_i + \varepsilon_{it}, \tag{1}$$

where  $i = 1, \dots, N, t = 1, \dots, T$  and  $\phi' = (\phi_1, \phi_2, \dots, \phi_T)$ , the  $\alpha_i$  and  $\varepsilon_{it}$  are independent and identically distributed (i.i.d.) random variables with mean 0 and variances  $\sigma_\alpha^2$  and  $\sigma_\varepsilon^2$ , respectively, which are mutually independent. Throughout this paper we assumed that the parameters  $\sigma_\alpha^2$  and  $\sigma_\varepsilon^2$  are a priori known, for simplicity. If these parameters are unknown, the consistent estimators can be used in their place and the same results hold true asymptotically [18]. Important special cases are the determination of the optimal design if the parameter of interest is the period mean  $\beta_p$ , if the parameter of interest is the change in two subsequent period means  $\beta_t - \beta_{t-1}$ , or if the parameter of interest is the overall average of the period means  $\sum_{t=1}^T \beta_t$ .

We denote the sample size in each wave by  $N$  and the proportion of cross-sections in all samples by  $\lambda N$ , while the remaining  $(1-\lambda)N$  individuals will be re-interviewed every period. In order to determine the optimal value of  $\lambda$  (i.e., the proportion of cross-sections in all samples) we first derive the efficient estimator and its variance. It is well known that the estimator of  $\beta' = (\beta_1, \beta_2, \dots, \beta_T)$  in Eq (1), using only the information on individuals which are re-interviewed every period, is the best linear unbiased estimator and regarded as  $\hat{\beta}_p$  [11]. Analogously, the estimator based on the cross-section information only is also the best linear unbiased estimator

and regarded as  $\widehat{\beta}_{cs}$  [11], and that

$$\text{var}(\widehat{\beta}_t) = \frac{1}{J^2} \text{var}\left(\sum_{i=1}^J \alpha_i + \sum_{i=1}^J \varepsilon_{it}\right) = \frac{\sigma_\varepsilon^2 + \sigma_\alpha^2}{J}, \tag{2}$$

$$\text{cov}(\widehat{\beta}_t, \widehat{\beta}_s) = \frac{1}{J^2} \text{cov}\left(\sum_{i=1}^J \alpha_i + \sum_{i=1}^J \varepsilon_{it}, \sum_{i=1}^J \alpha_i + \sum_{i=1}^J \varepsilon_{is}\right) = \frac{\sigma_\alpha^2}{J}, (s \neq t) \tag{3}$$

where  $J$  denotes the number of observed individuals. Therefore

$$\text{var}(\widehat{\beta}_p) = \frac{1}{(1-\lambda)N} (\sigma_\varepsilon^2 I_T + \sigma_\alpha^2 l_T l_T') = \frac{1}{(1-\lambda)N} V_p, \tag{4}$$

$$\text{var}(\widehat{\beta}_{cs}) = \frac{1}{\lambda N} (\sigma_\varepsilon^2 + \sigma_\alpha^2) I_T = \frac{1}{\lambda N} V_{cs}, \tag{5}$$

where  $l_T = (1, \dots, 1)_{T \times 1}'$  and  $V_p = \sigma_\varepsilon^2 I_T + \sigma_\alpha^2 l_T l_T'$ ,  $V_{cs} = (\sigma_\varepsilon^2 + \sigma_\alpha^2) I_T$ .

Since  $\widehat{\beta}_p$  and  $\widehat{\beta}_{cs}$  are independent, based on the relative theory of two sample estimation [10, 11], the best linear unbiased estimator which uses all the samples is given by

$$\widehat{\beta} = \left[ (1-\lambda)N V_p^{-1} + \lambda N V_{cs}^{-1} \right]^{-1} \left[ (1-\lambda)N V_p^{-1} \widehat{\beta}_p + \lambda N V_{cs}^{-1} \widehat{\beta}_{cs} \right]. \tag{6}$$

For

$$\text{var} \left[ (1-\lambda)N V_p^{-1} \widehat{\beta}_p + \lambda N V_{cs}^{-1} \widehat{\beta}_{cs} \right] = (1-\lambda)N V_p^{-1} + \lambda N V_{cs}^{-1}, \tag{7}$$

it is easily verified that

$$\text{var}(\phi' \widehat{\beta}) = N^{-1} \phi' \left\{ \lambda V_{cs}^{-1} + (1-\lambda) V_p^{-1} \right\}^{-1} \phi. \tag{8}$$

Consequently, the efficiency of split panel design could be transformed into the following nonlinear optimization by minimizing the variance of the best unbiased estimator of  $\phi' \widehat{\beta}$

$$\min_{N, \lambda} N^{-1} \phi' \left\{ \lambda V_{cs}^{-1} + (1-\lambda) V_p^{-1} \right\}^{-1} \phi. \tag{9}$$

In order to obtain the optimal solution for  $\lambda$  from Eq (9), the main steps are discussed in the derivation of the manageable expression for the variance of the best unbiased estimator  $\phi' \widehat{\beta}$  in Eq (1). First,  $V_p^{-1}$  and  $V_{cs}^{-1}$  can be written as

$$V_p^{-1} = \frac{1}{(\sigma_\varepsilon^2 + \sigma_\alpha^2)} V_{p0}^{-1} \tag{10}$$

and

$$V_{cs}^{-1} = \frac{1}{(\sigma_\varepsilon^2 + \sigma_\alpha^2)} I_T, \tag{11}$$

where

$$V_{p_0}^{-1} = \begin{bmatrix} \frac{(2-T)\rho - 1}{(\rho - 1)[1 + (T - 1)\rho]} & \frac{\rho}{(\rho - 1)[1 + (T - 1)\rho]} & \cdots & \frac{\rho}{(\rho - 1)[1 + (T - 1)\rho]} \\ \frac{\rho}{(\rho - 1)[1 + (T - 1)\rho]} & \frac{1 + (T - 2)\rho}{(1 - \rho)[1 + (T - 1)\rho]} & \cdots & \frac{\rho}{(\rho - 1)[1 + (T - 1)\rho]} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\rho}{(\rho - 1)[1 + (T - 1)\rho]} & \frac{\rho}{(\rho - 1)[1 + (T - 1)\rho]} & \cdots & \frac{1 + (T - 2)\rho}{(1 - \rho)[1 + (T - 1)\rho]} \end{bmatrix} \quad (12)$$

and

$$\rho = \frac{\sigma_z^2}{(\sigma_e^2 + \sigma_z^2)}. \quad (13)$$

Since  $V_{cs}^{-1}$  is a constant and multiple of the identity matrix, and  $V_p^{-1}$  is symmetric, there exists an orthogonal matrix  $Q$  such that  $Q^T V_{p_0}^{-1} Q = D$  and  $Q^T V_{cs}^{-1} Q = \frac{1}{(\sigma_e^2 + \sigma_z^2)} I_T$ , where  $D$  is a diagonal matrix and written as

$$D = \begin{bmatrix} \frac{1}{1 + (T - 1)\rho} & 0 & \cdots & 0 \\ 0 & \frac{1}{1 - \rho} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \frac{1}{1 - \rho} \end{bmatrix}, \quad (14)$$

and the orthogonal matrix  $Q$  can be written as

$$Q = \begin{bmatrix} \frac{-1}{\sqrt{T}} & \frac{-1}{\sqrt{T}} & \cdots & \frac{-1}{\sqrt{T}} \\ \frac{-1}{\sqrt{T}} & \frac{1 + \sqrt{T} - T}{T - \sqrt{T}} & \cdots & \frac{1}{T - \sqrt{T}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-1}{\sqrt{T}} & \frac{1}{T - \sqrt{T}} & \cdots & \frac{1 + \sqrt{T} - T}{T - \sqrt{T}} \end{bmatrix}. \quad (15)$$

The proof of constructing the orthogonal matrix  $Q$  is presented in the Appendix A1.

As such, the variance of the best unbiased estimator of  $\phi' \hat{\beta}$  using all the samples is written as

$$\text{var}(\phi' \hat{\beta}) = N^{-1}(\sigma_z^2 + \sigma_e^2) \phi' Q' \{ \lambda I + (1 - \lambda) D \}^{-1} Q \phi, \quad (16)$$

where

$$\{\lambda I + (1 - \lambda)D\}^{-1} = \begin{bmatrix} \frac{1 + (T - 1)\rho}{1 + (T - 1)\rho\lambda} & 0 & \dots & 0 \\ 0 & \frac{1 - \rho}{1 - \lambda\rho} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \frac{1 - \rho}{1 - \lambda\rho} \end{bmatrix}. \tag{17}$$

We denote  $\phi'Q = (\delta_1, \dots, \delta_T) = \delta'$  to obtain the simple expression of Eq (16), and rewrite it as

$$\text{var}(\phi'\widehat{\beta}) = N^{-1}(\sigma_x^2 + \sigma_\varepsilon^2)\left(\frac{1 + (T - 1)\rho}{1 + (T - 1)\lambda\rho} \delta_1^2 + \sum_{t=2}^T \left(\frac{1 - \rho}{1 - \lambda\rho}\right) \delta_t^2\right). \tag{18}$$

Consequently, the nonlinear optimization by minimizing the variance of the best unbiased estimator of  $\phi'\widehat{\beta}$  can be rewritten as

$$\min_{N, \lambda} N^{-1}(\sigma_x^2 + \sigma_\varepsilon^2)\left(\frac{1 + (T - 1)\rho}{1 + (T - 1)\lambda\rho} \delta_1^2 + \sum_{t=2}^T \left(\frac{1 - \rho}{1 - \lambda\rho}\right) \delta_t^2\right). \tag{19}$$

### Split panel design efficiency irrespective of the parameters of interest and budget

By considering the linear combinations of vector  $\beta$ , we can then easily adapt the results to an individual element, difference of elements, or overall average. As such, in this section, we will derive a theorem for the split panel design efficiency, irrespective of the parameters of interest and budget, using Eq (19).

**Theorem 1** Pure panel ( $\lambda = 0$ ) will minimize the variance of the best unbiased estimator of  $\phi'\widehat{\beta}$ , irrespective of the choice of  $\phi$ ,

$$\text{if } \frac{\left(\sum_{t=1}^T \phi_t\right)^2}{\sum_{t=1}^T \phi_t^2} < \frac{1 - \rho}{1 + T\rho - 2\rho}; \tag{20}$$

pure series of cross-sections ( $\lambda = 1$ ) will minimize the variance of the best unbiased estimator of  $\phi'\widehat{\beta}$ , irrespective of the choice of  $\phi$ ,

$$\text{if } \frac{\left(\sum_{t=1}^T \phi_t\right)^2}{\sum_{t=1}^T \phi_t^2} > \max\left\{\frac{1 - \rho}{1 + T\rho - 2\rho}, 1 + T\rho - \rho\right\}; \tag{21}$$

split panel ( $\lambda = k_r$ ) will minimize the variance of the best unbiased estimator of  $\phi'\widehat{\beta}$ ,

irrespective of the choice of  $\phi$ ,

$$\text{if } \frac{1 - \rho}{1 + T\rho - 2\rho} < \frac{\left(\sum_{t=1}^T \phi_t\right)^2}{\sum_{t=1}^T \phi_t^2} < \min\{(1 - \rho)(T - 1), 1 + T\rho - \rho\}; \tag{22}$$

split panel ( $\lambda = k_l$ ) will minimize the variance of the best unbiased estimator of  $\phi' \hat{\beta}$ . irrespective of the choice of  $\phi$ ,

$$\text{if } (1 - \rho)(T - 1) < \frac{\left(\sum_{t=1}^T \phi_t\right)^2}{\sum_{t=1}^T \phi_t^2} < 1 + T\rho - \rho; \tag{23}$$

split panel ( $\lambda = \lambda_0$ ) will minimize the variance of the best unbiased estimator of  $\phi' \hat{\beta}$ , irrespective of the choice of  $\phi$ ,

$$\text{if } \frac{\left(\sum_{t=1}^T \phi_t\right)^2}{\sum_{t=1}^T \phi_t^2} = (1 - \rho)(T - 1), \tag{24}$$

where  $k_l$  is the left root of  $k(\lambda)$ ;  $k_r$  is the right root of  $k(\lambda)$ ,

$$k(\lambda) = (1 - T)\rho A(1 - \lambda\rho)^2 + B\rho(1 + (T - 1)\lambda\rho)^2. \tag{25}$$

If

$$M\left(-\frac{c}{b}\right) < M(1), \lambda_0 = -\frac{c}{b}, \tag{26}$$

if

$$M\left(-\frac{c}{b}\right) > M(1), \lambda_0 = 1, \tag{27}$$

And

$$b = 2\rho^2(T - 1) \left[ (1 - \rho) \sum_{t=1}^T \varphi_t^2 + \rho \left( \sum_{t=1}^T \varphi_t \right)^2 \right], \tag{28}$$

$$c = (1 - \rho)\rho \sum_{t=1}^T \varphi_t^2 + \rho(2\rho - 1 - \rho T) \left( \sum_{t=1}^T \varphi_t \right)^2. \tag{29}$$

$$M(\lambda) = \frac{(1 - \lambda)\rho}{(1 + (T - 1)\lambda\rho)(1 - \lambda\rho)} \left( \sum_{t=1}^T \varphi_t \right)^2 + \frac{1 - \rho}{1 - \lambda\rho} \sum_{t=1}^T \varphi_t^2. \tag{30}$$

The proof of theorem 1 is presented in Appendix 1B.

From Theorem 1, it can be easily checked that  $\widehat{\beta}_t - \widehat{\beta}_{t-1}$  has the smallest variance if a pure panel ( $\lambda = 0$ ) is used. Likewise, a pure series of cross-sections ( $\lambda = 1$ ) will be optimal if the overall average of period means is to be estimated.

In order to illustrate that the split panel design will be preferable to pure panel or pure series of cross-sections design in most cases, and how much efficiency will be lost if a suboptimal choice is made when the period mean  $\beta_t$  is the parameter of interest, we present in Table 1 the relative efficiency of the estimator based on the split panel to an estimator based on a pure series of cross-sections or pure panel (pure series of cross-sections and pure panel yield equally efficient estimators in this case). Similar to [18], we assume the observation period  $T = 3, 6, 12, 20$ , the proportion of the component of variance  $\rho = 0.3, 0.6, 0.9$ , and the proportions of a series of cross-sections in split panel  $\lambda = 1/2, 1/3, 1/4, 1/8, 1/12$ .

### Split panel design efficiency with budget constraint and main algorithm

As opposed to the previous section, where we have analyzed split panel design efficiency, irrespective of the parameters of interest and budget, in this section, split panel design efficiency with a given budget is considered. Let  $p_1$  denote the average cost of observing every individual in cross-sections and  $p_2$  the average cost of observing every individual in panels. The cost of a cross-sectional survey is 30% to 70% higher than an additional wave of the panel study of income dynamics, as shown by Duncan et al. [19]. Therefore, we obtain  $0.6 < \frac{p_2}{p_1} < 0.8$ . If there is a budget,  $C$ , for all the periods, we can obtain the constrained nonlinear integer optimization (P1), as follows

$$\min z(\lambda, N) = N^{-1} \phi' \{ \lambda V_{cs}^{-1} + (1 - \lambda) V_p^{-1} \}^{-1} \phi, \tag{31}$$

$$s.t. \lambda N p_1 + (1 - \lambda) N p_2 \leq C^*, \tag{32}$$

**Table 1. The relative efficiency compared to pure cross-sections (or pure panel) for the estimator  $\widehat{\beta}_t$ .**

	$\rho$	$T = 3$	$T = 6$	$T = 12$	$T = 20$
$\lambda = 1/2$	0.3	0.9593	0.9244	0.8901	0.8694
	0.6	0.8393	0.7429	0.6711	0.6354
	0.9	0.6124	0.4336	0.3193	0.2675
$\lambda = 1/3$	0.3	0.9630	0.9259	0.8836	0.8544
	0.6	0.8571	0.7500	0.6563	0.6042
	0.9	0.6786	0.4857	0.3422	0.2708
$\lambda = 1/4$	0.3	0.9683	0.9337	0.8900	0.8571
	0.6	0.8778	0.7731	0.6704	0.6081
	0.9	0.7297	0.5389	0.3797	0.2941
$\lambda = 1/8$	0.3	0.9810	0.9569	0.9204	0.8865
	0.6	0.9260	0.8452	0.7434	0.6665
	0.9	0.8370	0.6806	0.5092	0.3955
$\lambda = 1/12$	0.3	0.9866	0.9687	0.9392	0.9092
	0.6	0.9474	0.8842	0.7946	0.7179
	0.9	0.8837	0.7568	0.5968	0.4759

doi:10.1371/journal.pone.0154913.t001

$$0 \leq \lambda \leq 1, \tag{33}$$

$$\lambda N, N \in N^*, \tag{34}$$

where  $C^* = \frac{c}{T}$ .

Applying Eq (19) and  $\lambda N = x, N = y$ , we obtain the constrained nonlinear integer optimization (P2):

$$\min_{x,y} z(x, y) = \sigma^2 \cdot \left( \frac{1 + (T - 1)\rho}{y + (T - 1)x\rho} \delta_1^2 + \sum_{i=2}^T \left( \frac{1 - \rho}{y - x\rho} \right) \delta_i^2 \right), \tag{35}$$

$$\text{s.t. } x(p_1 - p_2) + yp_2 \leq C^*, \tag{36}$$

$$0 \leq x \leq y, \tag{37}$$

$$x, y \in N^*, \tag{38}$$

where  $C^* = \frac{c}{T}$ .

Eq (35) is the objective function that minimizes the variance of the best linear unbiased estimator of linear combinations of the period means, while Eq (36) satisfies the constraint of a given budget.

**Algorithm design.** In section 4, we transformed the efficiency of split panel design into the constrained nonlinear integer optimization (P2), which is, nonetheless, difficult to solve with the current mature optimal algorithms. The simulated annealing algorithm has the advantages of guaranteeing global optimization, selecting the initial solution randomly, while being simple and practical. However, when it is used to solve the constrained nonlinear integer optimization associated with the efficiency of split panel design, given a budget, it is difficult to combine the parameters, such as the inner iteration number, the initial temperature, and the temperature decrease rate, in order to get the best performance of the algorithm. Consequently, in this paper, we design an efficient algorithm to solve the constrained nonlinear integer optimization associated with split panel design efficiency of given a budget, based on the simulated annealing algorithm.

The steps of the simulated annealing algorithm designed to solve (P2) are given as follows:

1. Choose the initial integer solution  $x_0 \in D$  and the initial temperature value  $T_0 > 0$ , where  $D$  is a feasible region formed by Eqs (4) and (5); calculate  $f(x_0)$  and let

$$X_0 = x_0, X_{\min} = X^0, f_{\min} = f(x^0), K = 0, \tag{39}$$

2. Randomly generate the integer vector

$$Z^K = (z_1^K, \dots, z_n^K), \tag{40}$$

where

$$z_i^K = \left\langle \text{sign}(U_i) \cdot T_k \cdot \left( \frac{1}{|U_i|^m} - 1 \right) \right\rangle, i = 1, 2, \dots, n, \tag{41}$$

and  $z_i^K$  is the  $i$ th component of the random vector  $Z^K$ ;  $U_1, U_2, \dots, U_n$  is a group of random

variables distributed uniformly over  $[-1,1]$ , which are independent each other;  $sign(\cdot)$  is the sign function; and  $\langle \cdot \rangle$  is the symbol of rounding numbers.

- Use the current iteration point  $x^K$  and the random vector  $Z^K$  to generate a new iteration point  $Y^K$  that satisfies  $Y^K = X^K + Z^K$ . If  $Y^K \in D$ , the next step is carried on, and if  $Y^K \notin D$ ,  $Y^K$  is calculated by

$$Y^K = X^K + \left\langle \left(\frac{-1}{2}\right)^l \cdot Z^K \right\rangle \tag{42}$$

until  $Y^K \in D$  and to the next step, where  $l = 1, 2, \dots, N_1$ . If  $Y^K \notin D$  in the  $N_1$  steps, let  $Y^K = X^K$  and move to the next step.

- Generate a random number  $\eta$  distributed uniformly over  $[0,1]$  and calculate

$$P_a(Y^K | X^K, T_K) = \min \left\{ 1, \exp\left(\frac{f(X^K) - f(Y^K)}{\beta T_K}\right) \right\} \tag{43}$$

using the current iteration point  $X^K$  and a new iteration point,  $Y^K$ .

- If

$$\eta \leq P_a(Y^K | X^K, T_K), \tag{44}$$

let

$$X^{K+1} = Y^K, f(X^{K+1}) = f(Y^K) \tag{45}$$

or let

$$X^{K+1} = X^K, f(X^{K+1}) = f(X^K). \tag{46}$$

- If

$$f(X^{K+1}) < f_{\min}, \tag{47}$$

let

$$X_{\min} = X^{K+1}, f_{\min} = f(X^{K+1}). \tag{48}$$

- If the stopping criterion satisfies

$$|f(X^{K+1}) - f(X^K)| < \varepsilon, \tag{49}$$

stop calculating and regard  $X_{\min}$  and  $f_{\min}$  as the approximate global optimal solution and the corresponding optimal value, respectively. If not, move to the next step.

- Generate a new temperature  $T_{K+1}$  by using the given renewed function of temperature as follows:

$$T_{K+1} = \frac{T_0}{(K+1)^m}, K = 0, 1, 2, \dots \tag{50}$$

and let  $K = K+1$  and shift to the second step.

The detailed design process of simulated annealing algorithm is presented in Appendix 1C

## Results and Discussion

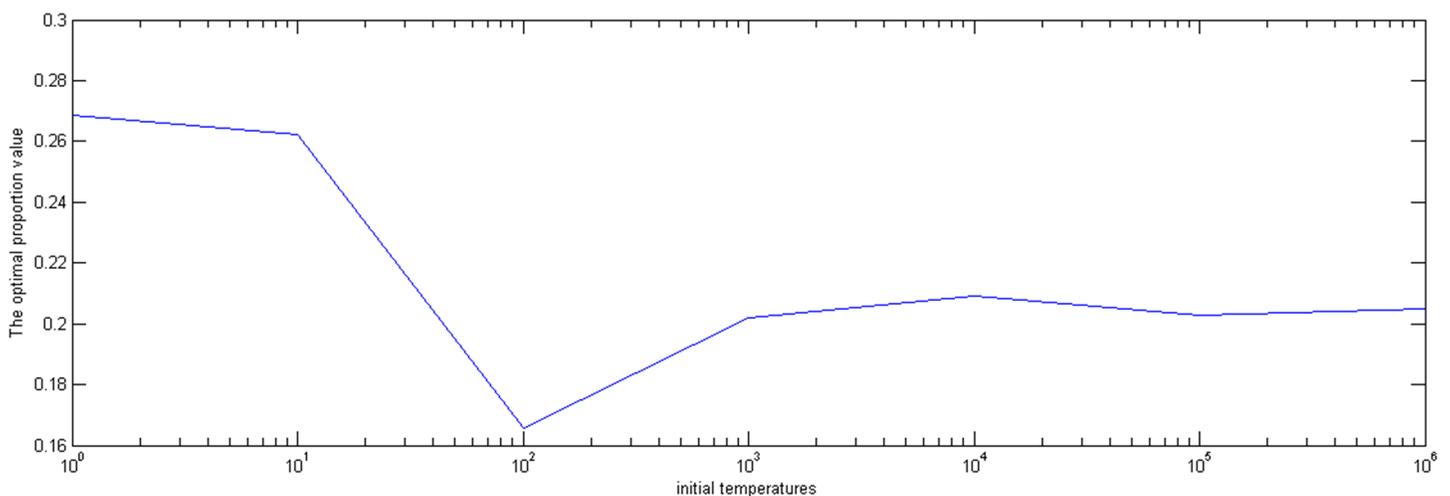
The example used in this study is the monthly consumer expenditure on food in 1985, in the Netherlands, which is modeled using [Model \(1\)](#) and the so-called expenditure index panel conducted by Infomart, a marketing research agency [11]. We restrict analysis to  $\xi_1 = \xi_2 = \dots = \xi_{12}$  (annual average). The maximum likelihood estimate of  $\rho$  in [Eq \(1\)](#) for food is 0.76, with standard error 0.005 [11]. From [19], the survey cost was estimated to be roughly USD 513,000. The average cost of observing every individual in cross-section  $p_1$  and the average cost of observing every individual in panels  $p_2$  were estimated to be roughly USD 125 and USD 75, respectively. The following results are obtained using MATLAB.

### OATD method

The benchmarking parameter combinations of the algorithm designed in this paper are set as follows: the inner iteration number  $B = 2000$ , the initial temperature  $E_0 = 10000$ , the temperature decrease rate  $m = 0.75$  and the termination temperature  $\varepsilon = 0.0001$ . Subsequently, we analyze the effects of these parameters.

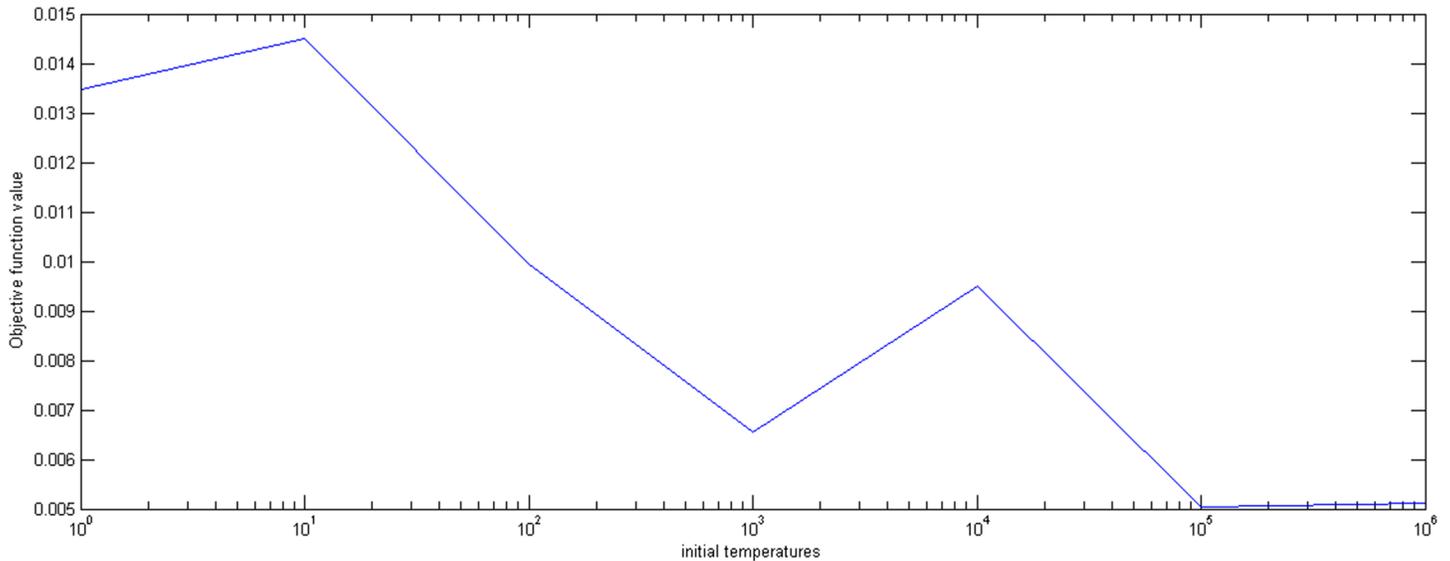
First, we set the inner iteration number  $B$ , the temperature decrease rate  $m$ , the termination temperature  $\varepsilon$ , and the initial temperature  $E_0$  is changed from 1 to 1,000,000. The optimal proportion values and the corresponding objective values are shown in [Fig 1](#). From [Fig 2](#), the objective values fluctuate between 0.005 and 0.015 and, when the initial temperature is more than 100,000, the optimal proportion values and the corresponding objective values can converge to the optimal values of 0.2 and 0.005, respectively. Therefore, the initial temperature can be chosen between 100,000 and 1,000,000.

Second, we set the initial temperature  $E_0$ , the temperature decrease rate  $m$ , the termination temperature  $\varepsilon$ , and the inner iteration number  $B = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$  are changed from 500 to 3,000 to ensure that the algorithm reaches the balanced state. The optimal proportion values and the corresponding objective values are shown in [Fig 3](#). From [Fig 3](#), the higher the inner iteration number is, the more easily the algorithm moves from the local optimal value and converges to the global optimal value. Conversely, the higher the inner iteration number is, the longer the implementation time. In this study, the implementation time based on the set of parameters is in 10 minutes. As such, we do not consider that the inner iteration numbers increases



**Fig 1. The effect of initial temperature on the optimal proportion value.**

doi:10.1371/journal.pone.0154913.g001

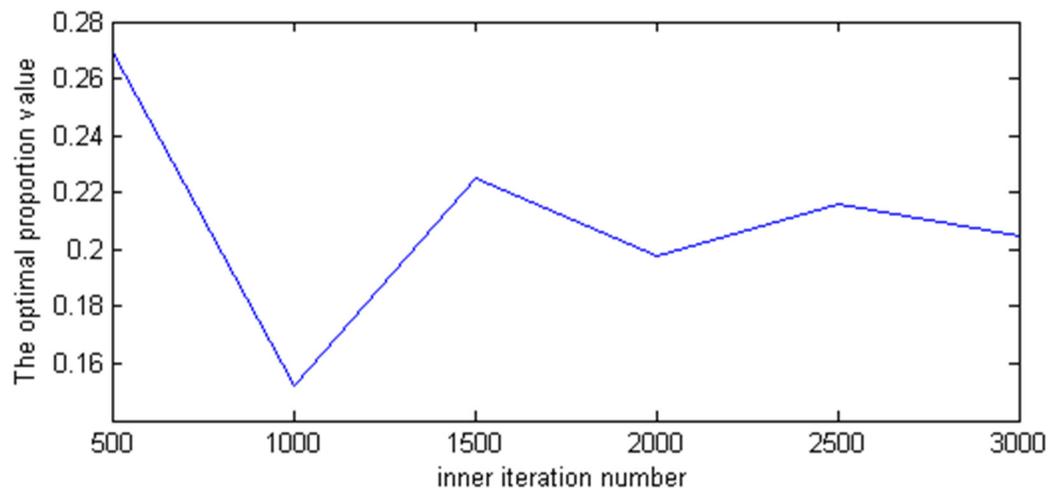


**Fig 2. The effect of initial temperature on the objective function value.**

doi:10.1371/journal.pone.0154913.g002

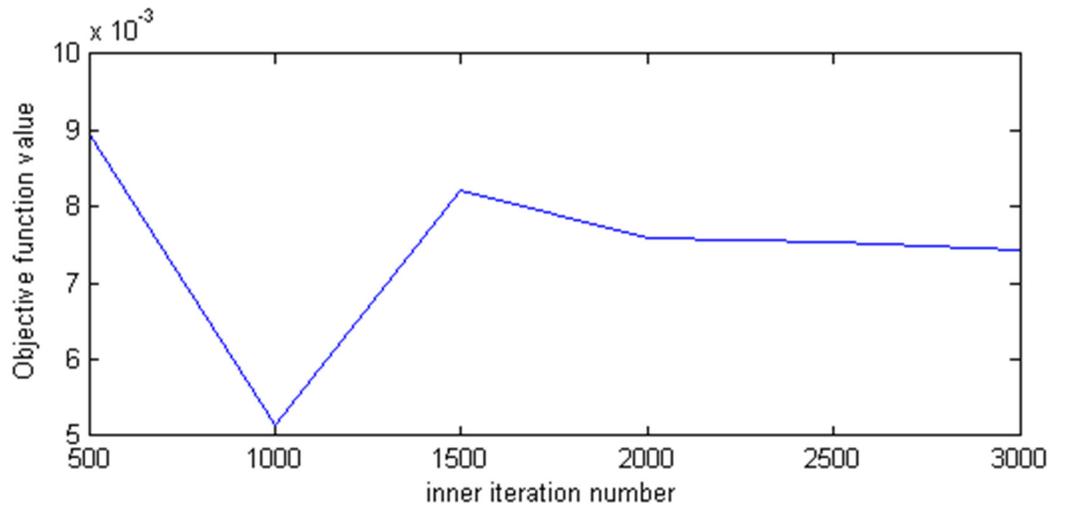
implementation time. From Fig 4, when the inner iteration number is more than 2,000, the objective values and the optimal proportion values converge to 0.007 and 0.2, respectively. Therefore, the inner iteration number can be chosen between 2,000 and 2,500.

Subsequently, we set the initial temperature  $E_0$ , the termination temperature  $\epsilon$ , the inner iteration number  $B$ , and the temperature decrease rate  $m$  is changed from 0.45 to 0.9. The optimal proportion values and the corresponding objective values are shown in Fig 5. If the temperature decrease rate is less than 0.6, the algorithm falls into the local optimal value. When the temperature decrease rate is more than 0.6, the algorithm is above the local optimal value and converges to the global optimal value. As is shown in Figs 5 and 6, the optimal proportion values and the corresponding objective values converge to 0.2 and 0.0046, respectively. The temperature decrease rate determines the searching space and the larger the temperature decrease



**Fig 3. The effect of inner iteration number on the optimal proportion value.**

doi:10.1371/journal.pone.0154913.g003

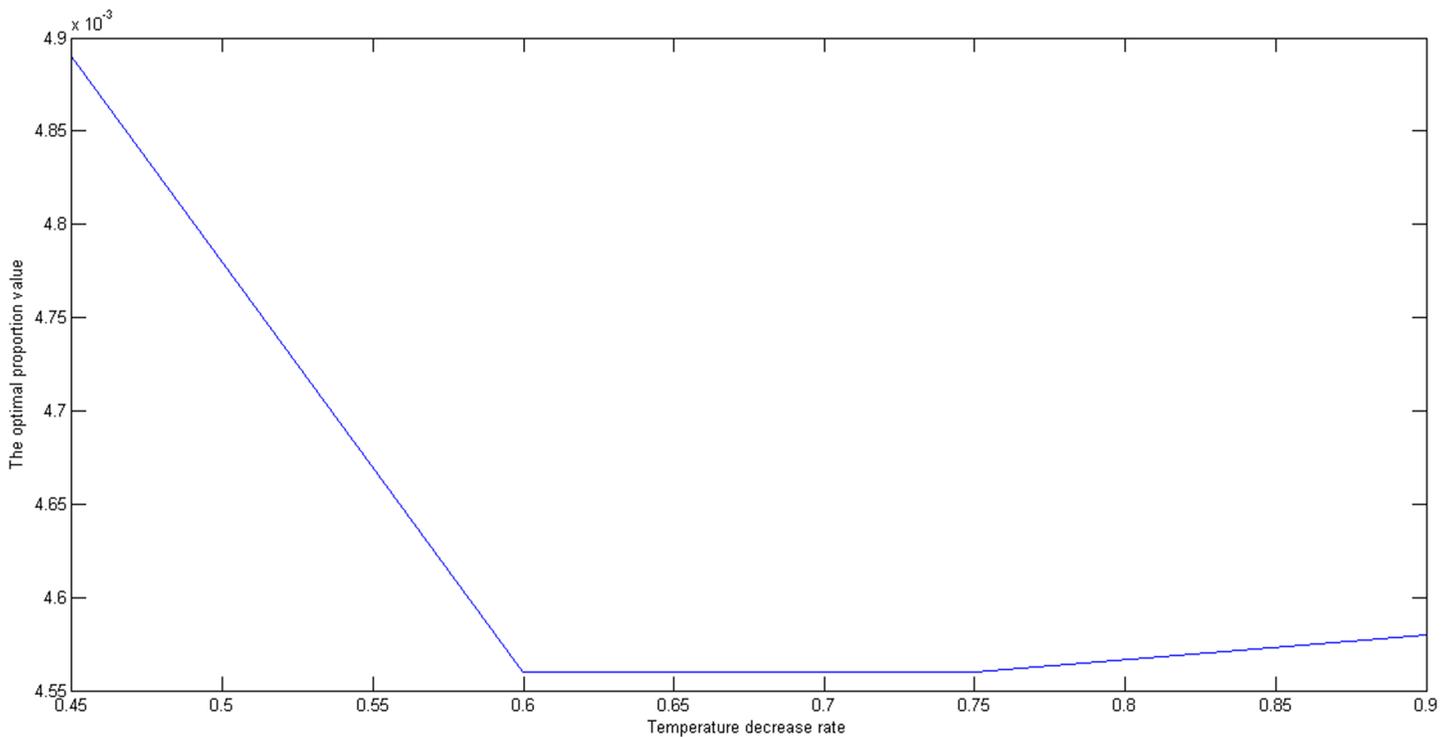


**Fig 4. The effect of inner iteration number on the objective function value.**

doi:10.1371/journal.pone.0154913.g004

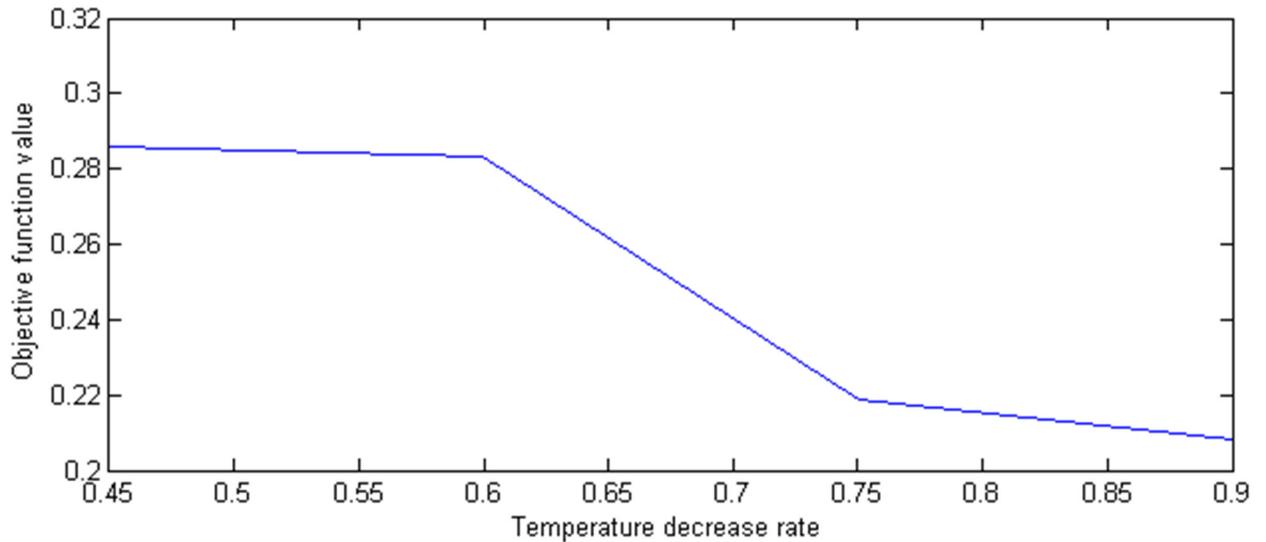
rate is, the greater the searching space. Consequently, we can choose the temperature decrease rate between 0.75 and 0.9.

Finally, we set the initial temperature  $E_0$ , the temperature decrease rate  $m$ , the inner iteration number  $B$ , and the termination temperature  $\varepsilon$  is changed from 0.00001 to 1. Figs 7 and 8 show that when the termination temperature  $\varepsilon$  is near 0.0001, the optimal proportion values and the corresponding objective values converge to 0.2 and 0.0043, respectively. The lower the



**Fig 5. The effect of temperature decrease on the optimal proportion value.**

doi:10.1371/journal.pone.0154913.g005



**Fig 6. The effect of temperature decrease on the objective function value.**

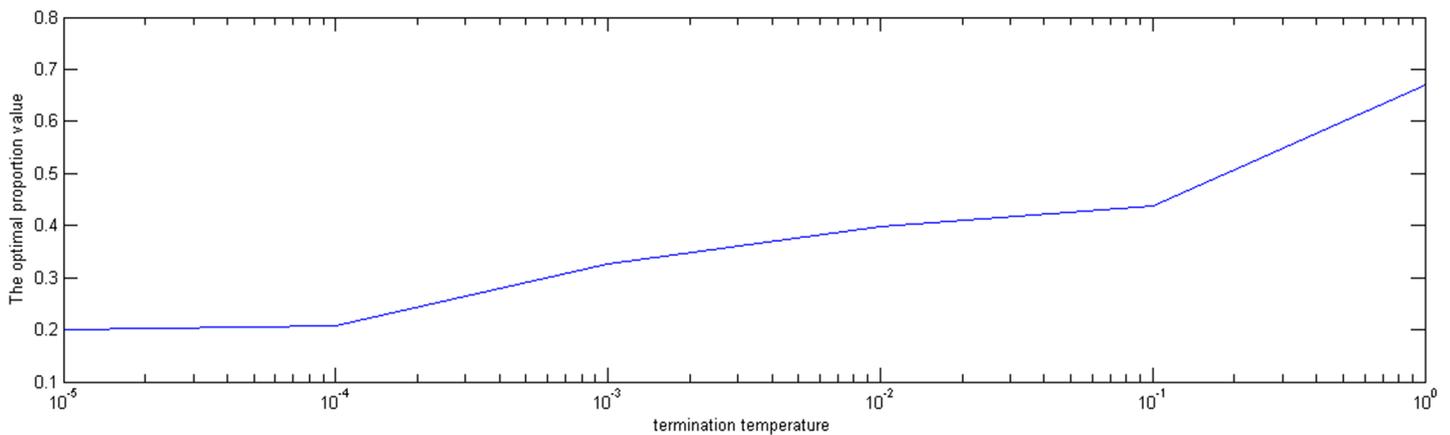
doi:10.1371/journal.pone.0154913.g006

termination temperature is, the more adequate the time to converge to the optimal value. Therefore, the termination temperature can be chosen between 0.0001 and 0.00001.

### FD method

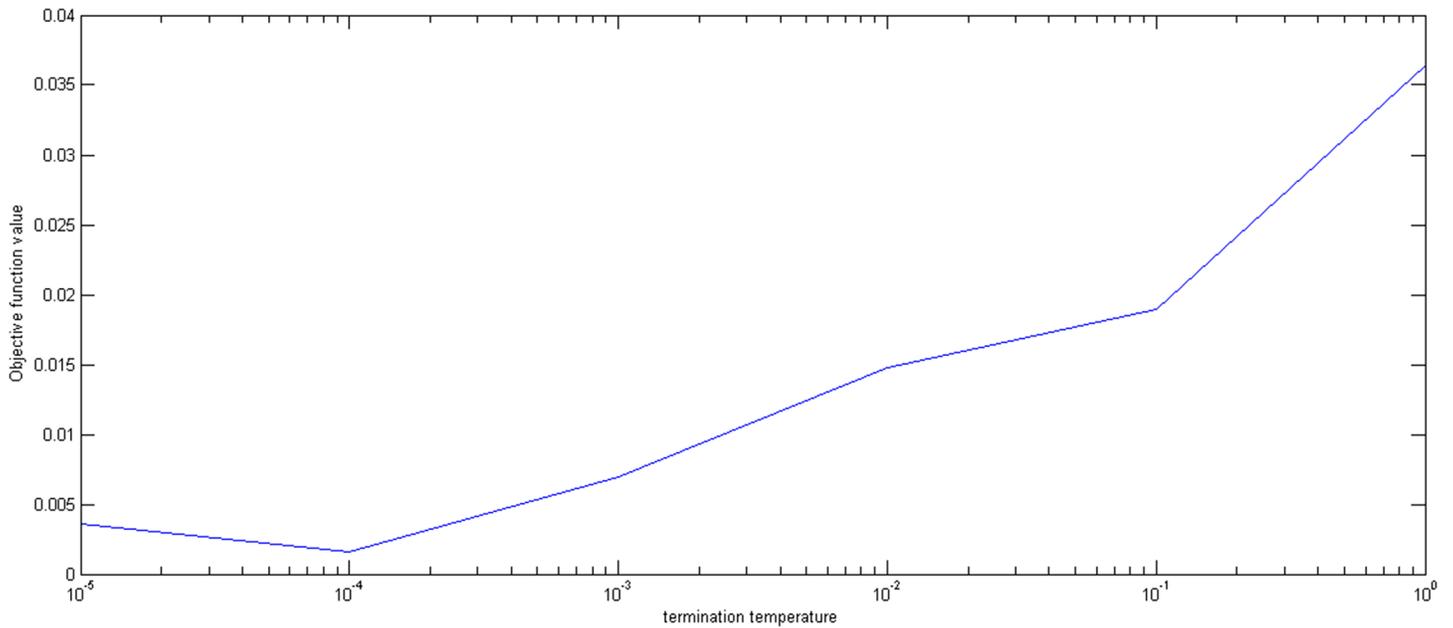
According to the above discussion, we separate the algorithm designed in this paper into two cases to analyze, the best and worst case. Under the two different cases, the fields of parameters based on the above OATD analysis can be separated into two levels (+ and -) as shown in [Table 2](#).

For each case, the temperature decrease rate  $m$ , the inner iteration number  $B$  and the termination temperature  $\epsilon$  are chosen randomly from their two parameter levels, and there are six values in all. Under each parameter combination, we take the optimal value 10 times the average value. The initial temperature is considered constant as compared to the whole field, since the local optimal value because of the initial temperature is very small, and the resultant errors



**Fig 7. The effect of termination temperature on the optimal proportion value.**

doi:10.1371/journal.pone.0154913.g007



**Fig 8. The effect of termination temperature on the objective function value.**

doi:10.1371/journal.pone.0154913.g008

from the initial temperature are smaller than that from other parameters. Comparing to other parameters, the initial temperature has a low quantitative influence on the algorithm designed in this paper, and will not be analyzed as such.

It is found as per Table 3 that the effects of the temperature decrease rate  $m$  and the inner iteration number  $B$  are clear, meaning that the choice of the two parameters decides whether the algorithm can obtain the optimal value. In different cases, the temperature decrease rate and the inner iteration number have different effects on the optimal value. In the best case, the inner iteration number is large enough to guarantee the solutions are stable and, if it continues to increase, the solutions do not improve much. In this case, the adjustment of temperature decrease rate can continue to narrow the neighborhood range for more convergent results. Therefore, the temperature decrease rate influences the best case. In the worst case, the inner iteration number has a significant impact, since the lower inner iteration number causes a wider searching range, which makes the process of searching far from the optimal solution. This time, the result is easier to move to the local optimal solution with the adjustment of the temperature decrease. Furthermore, Table 3 shows that the combination effect of the temperature decrease rate and the inner iteration number is also clear.

**Table 2. Parameters levels.**

Case	Parameter level	$B$	$\epsilon$	$m$
The best	+	[100000,1000000]	[0.00001,0.0001]	[0.85,0.90]
	-	[10000,100000]	[0.0001,0.001]	[0.80,0.85]
The worst	+	[1000,10000]	[0.001,0.1]	[0.75,0.80]
	-	[100,1000]	[0.01,0.1]	[0.60,0.75]

doi:10.1371/journal.pone.0154913.t002

Table 3. Index effects on numerical results.

Parameters level							The best	The worse
Run	<i>B</i>	<i>m</i>	$\epsilon$	( <i>B</i> , $\epsilon$ )	( <i>B</i> , <i>m</i> )	( <i>m</i> , $\epsilon$ )	<i>y</i>	<i>y</i>
1	-	-	-	+	+	+	0.2100	0.2624
2	+	-	-	-	-	+	0.1958	0.2518
3	-	+	-	-	+	-	0.2020	0.2280
4	+	+	-	+	-	-	0.2190	0.2308
5	-	-	+	+	-	-	0.2030	0.2468
6	+	-	+	-	+	-	0.1993	0.2217
7	-	+	+	-	-	+	0.2315	0.2970
8	+	+	+	+	+	+	0.2301	0.2860
	-0.0260	-0.0393	-0.0120	0.0221	-0.0538	0.0112	The best	Index
	-0.0310	-0.0460	-0.0078	0.0230	-0.0760	0.0218	The worse	

doi:10.1371/journal.pone.0154913.t003

### Conclusions

In this paper, we discuss how to determine the optimal proportion of a series of cross-sections in all samples to minimize survey design efficiency in the analysis-of-variance model, which can be applied directly in sampling. First, we derive a theorem for choosing the optimal proportion of a series of cross-sections in all samples, irrespective of the parameters of interest and budget. In addition, our results show that, compared to a pure series of cross-sections or pure panel, the gains from choosing split panel can be substantial. Second, the efficiency of split panel design given a budget is considered and an efficient algorithm is designed to solve the constrained nonlinear integer optimization associated with the efficiency of survey designs on a budget. We further apply OATD and FD methods to analyze and compare the quantitative influence of different selections of parameters in the implementation of the algorithm with an empirical example concerning monthly consumer expenditure on food in 1985, in the Netherlands, and obtain the efficient ranges of the algorithm parameters to ensure a good solution.

For further research, we will extend the results to a more general analysis of the covariance model and derive the expressions for the variances of efficient parameter estimators. At the same time, other algorithms can be to solve the new nonlinear programming from the optimal split design in the analysis of the covariance model.

### Appendix 1

#### Appendix 1A. Proof of constructing the orthogonal matrix Q

Proof:

$$\begin{aligned}
 Q^T Q &= \left[ \frac{1}{(\zeta, \zeta) - (\zeta, \beta)} \right]^2 \{ (\zeta - \beta)(\zeta - \beta)^T (\zeta - \beta)(\zeta - \beta)^T \\
 &\quad - 2[(\zeta, \zeta) - (\zeta, \beta)](\zeta - \beta)(\zeta - \beta)^T + [(\zeta, \zeta) - (\zeta, \beta)]^2 \} = I
 \end{aligned}
 \tag{51}$$

Therefore, matrix Q is an orthogonal matrix.

$\frac{1}{1-\rho}$  is the n-repeated eigenvalue of real symmetric matrix  $V_{p_0}^{-1}$ , so

$$\text{rank}\left(\frac{1}{1-\rho} I - V_{p_0}^{-1}\right) = 1.
 \tag{52}$$

Let  $\zeta^T$  denote the nonzero row vector of  $\frac{1}{1-\rho}I - V_{p_0}^{-1}$ , so  $(\frac{1}{1-\rho}I - V_{p_0}^{-1})X = 0$  and  $\zeta^T X = 0$  have the same solutions. Let  $B_1$  denote the matrix removing the first column of matrix  $Q$ ,

$$\zeta^T = (a_1, a_2, \dots, a_n) \tag{53}$$

and

$$\zeta_1^T = (a_2, a_3, \dots, a_n). \tag{54}$$

Subsequently,

$$B_1 = \frac{1}{(\zeta, \zeta) - (\zeta, \beta)} \{(\zeta - \beta)\zeta_1^T - [(\zeta, \zeta) - (\zeta, \beta)]\zeta^T\} \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & & & 1 \end{bmatrix}, \tag{55}$$

$$\begin{aligned} \zeta^T B_1 &= \frac{1}{(\zeta, \zeta) - (\zeta, \beta)} \{ \zeta^T (\zeta - \beta)\zeta_1^T - [(\zeta, \zeta) - (\zeta, \beta)]\zeta^T \} \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & & & 1 \end{bmatrix} \\ &= \frac{1}{(\zeta, \zeta) - (\zeta, \beta)} \{ \zeta^T (\zeta - \beta)\zeta_1^T - [(\zeta, \zeta) - (\zeta, \beta)]\zeta^T \} \\ &= \frac{1}{(\zeta, \zeta) - (\zeta, \beta)} \{ \zeta^T (\zeta - \beta)\zeta_1^T - \zeta^T (\zeta - \beta)\zeta_1^T \} = 0_{1 \times n-1} \end{aligned} \tag{56}$$

As such, the bottom  $n-1$  column vectors of matrix  $Q$  are the solutions of  $\zeta^T X = 0$ . Matrix  $Q$  is an orthogonal matrix and it is impossible for all the column vectors of matrix  $Q$  to be zero, therefore, the bottom  $n-1$  column vectors of matrix  $Q$  are eigenvectors belonging to eigenvalue  $\frac{1}{1-\rho}$  of matrix  $V_{p_0}^{-1}$ . The  $n$  column vectors of the orthogonal matrix constitute a unit orthogonal vector group, and the bottom  $n-1$  column vectors are unit orthogonal vector group constituted of  $n-1$  eigenvectors belonging to eigenvalue  $\frac{1}{1-\rho}$  of matrix  $V_{p_0}^{-1}$ . The first column vector of matrix  $Q$  is a unit eigenvector belonging to eigenvalue  $\frac{1}{1+(T-1)\rho}$  of matrix  $V_{p_0}^{-1}$ . Moreover, different eigenvectors belonging to different eigenvalues are orthogonal, and the unit vector that is orthogonal with the bottom  $n-1$  column vectors of matrix  $Q$  is unit eigenvector belonging to eigenvalue  $\frac{1}{1+(T-1)\rho}$  of matrix  $V_{p_0}^{-1}$ . Additionally, matrix  $Q$  is an orthogonal matrix, and the first column vector of matrix  $Q$  is the unit eigenvector belonging to eigenvalue  $\frac{1}{1+(T-1)\rho}$  of matrix  $V_{p_0}^{-1}$ .

### Appendix 1B. Proof of theorem 1

Proof:

Let

$$M(\lambda) = \frac{1 + (T-1)\rho}{1 + (T-1)\lambda\rho} \delta_1^2 + \frac{1-\rho}{1-\lambda\rho} \sum_{t=2}^T \delta_t^2, \tag{57}$$

$$A = [1 + (T-1)\rho]\delta_1^2, \tag{58}$$

and

$$B = (1 - \rho) \sum_{t=2}^T \delta_t^2. \tag{59}$$

Subsequently,

$$M(\lambda) = \frac{A}{1 + (T - 1)\lambda\rho} + \frac{B}{1 - \lambda\rho}. \tag{60}$$

Therefore,

$$M'_\lambda(\lambda) = \frac{(1 - T)\rho A(1 - \lambda\rho)^2 + B\rho(1 + (T - 1)\lambda\rho)^2}{(1 + (T - 1)\lambda\rho)^2(1 - \lambda\rho)^2}. \tag{61}$$

Let

$$k(\lambda) = (1 - T)\rho A(1 - \lambda\rho)^2 + B\rho(1 + (T - 1)\lambda\rho)^2, \tag{62}$$

thus

$$k'_\lambda(\lambda) = 2\rho^2(T - 1)(1 - \lambda\rho)A + 2B\rho^2(T - 1)(1 + (T - 1)\lambda\rho), \tag{63}$$

when  $\lambda \in [0, 1], k'_\lambda \geq 0$ .

In detail,

$$k(\lambda) = ((1 - T)\rho^3 A + B(1 - T)^2 \rho^3) \lambda^2 + (2B\rho^2(T - 1) - 2(1 - T)\rho^2 A) \lambda + (1 - T)\rho A + B\rho. \tag{64}$$

Let

$$a = (1 - T)\rho^3 A + B(1 - T)^2 \rho^3, \tag{65}$$

$$b = 2B\rho^2(T - 1) - 2(1 - T)\rho^2 A, \tag{66}$$

$$c = (1 - T)\rho A + B\rho. \tag{67}$$

Subsequently,

$$k(\lambda) = a\lambda^2 + b\lambda + c. \tag{68}$$

Let  $\Delta_k = b^2 - 4ac$ , and as a result

$$\Delta_k = 4(1 - \rho) \sum_{t=2}^T \delta_t^2 (1 + (T - 1)\rho) \delta_1^2 \rho^4 (T - 1) T^2 \geq 0, (\lambda \in [0, 1]). \tag{69}$$

From the above results,

When  $\lambda \in [0, 1], c < 0$  and  $k(1) < 0$ ,  $M(1)$  is the minimum value of  $M(\lambda)$ .

When  $\lambda \in [0, 1]$  and  $c > 0$ ,  $M(0)$  is the minimum value of  $M(\lambda)$ .

When  $\lambda \in [0, 1], a < 0, c < 0$ , and  $k(1) > 0$ ,  $M(k_r)$  is the minimum value of  $M(\lambda)$ , where  $k_r$  is the right root of  $k(\lambda)$ .

When  $\lambda \in [0, 1], a > 0, c < 0$ , and  $k(1) > 0$ ,  $M(k_l)$  is the minimum value of  $M(\lambda)$ , where  $k_l$  is the left root of  $k(\lambda)$ .

When  $a = 0$ , and then  $c < 0$  and  $b < 0$ ,  $\min\{M(-\frac{c}{b}), M(1)\}$  is the minimum value of  $M(\lambda)$ .

### Appendix 1C. Detailed design process of proposed algorithm

Simulated annealing begins with an initial solution, and then randomly generates a neighboring solution or by using a pre-specified rule. It is the process when a state moves from the initial solution to a candidate solution in which the energy is minimized based on the Metropolis acceptance criterion. As such, we can accept the candidate solution based on the acceptance probability. We consider a number of conditions, and, subsequently, the steps of the proposed algorithm are shown.

In order to converge to the global optimal solution, we choose

$$\eta \leq P_a(Y^K|X^K, T_K) \tag{70}$$

as the acceptance criterion, which is called Metropolis criterion, where  $n$  is a random number distributed uniformly over  $[0,1]$  and the state acceptance function is denoted by

$$P_a(Y^K|X^K, T_K) = \min \left\{ 1, \exp\left(\frac{f(X^K) - f(Y^K)}{\beta T_K}\right) \right\}; \tag{71}$$

$f$  is the objective function of (P2);  $X^K$  and  $Y^K$  are the current iteration point and the new iteration point, respectively;  $T_K$  is the  $k$  time algorithm stage temperature, which is obtained from the cooling schedule presented in Eq (50);  $\beta$  is a positive constant. In order to guarantee the integer optimal solution, the new iteration point  $Y^K$  is generated by the following process:

$$Y^K = X^K + \left\langle \left(\frac{-1}{2}\right)^l \cdot Z^K \right\rangle, l = 1, 2, \dots, N_1, \tag{72}$$

where

$$z_i^K = \left\langle \text{sign}(U_i) \cdot T_K \cdot \left(\frac{1}{|U_i|^m} - 1\right) \right\rangle, i = 1, 2, \dots, n, \tag{73}$$

and  $z_i^K$  is the  $i$ th component of the random vector  $Z^K$ ;  $U_1, U_2, \dots, U_n$  is a group of random variables distributed uniformly over  $[-1,1]$ , which are independent of each other;  $\text{sign}(\cdot)$  is the sign function;  $\langle \cdot \rangle$  shows the symbol of rounding numbers.

The performance and convergence of simulated annealing are crucial, and are affected by the cooling schedule. If  $T$  decreases fast, a fast convergence can be obtained. However, the simulated annealing reaches the global optimal solution with difficulty if  $T$  decreases fast. Therefore, the following cooling schedule guarantees the global optimal solution by the above theory:

$$T_{K+1} = \frac{T_0}{(K + 1)^m}, K = 0, 1, 2 \dots, \tag{74}$$

where  $T_0$  is an initial temperature;  $m \geq 1$  and  $m$  is an integer constant, which determines the speed of the decreasing temperature. The choice of  $T_0$  may be crucial, the sophisticated techniques being discussed by Van Laarhoven and Aarts [12].

In conclusion, the stopping criterion for the simulated annealing algorithm is given by

$$|f(X^{K+1}) - f(X^K)| < \varepsilon, \tag{75}$$

where  $\varepsilon$  denotes any small number.

### Acknowledgments

I would like to extend my sincere gratitude to John Knight, Yan Xu and Ming Huang for their instructive advice and useful suggestions for my paper.

## Author Contributions

Conceived and designed the experiments: XL. Performed the experiments: XL. Analyzed the data: XL. Wrote the paper: XL. Contributed to update the language: WW HL.

## References

1. Binder D.A., Hidioglou M.A.. Sampling in time. Krishnaiah P.R. and Rao C.R., eds. 1988.
2. Kish L.. Data collection for details over space and time. *Statistical method and the improvement of data quality*, 1983; 28(2):73–84.
3. Kish L.. Timing of surveys for public policy. *Austral. J. Statist.* 1986; 28(1):1–12.
4. Rodríguez-Oreggia E.. Hurricanes and labor market outcomes: Evidence for Mexico. *Global environmental change*. 2013; 23(2):351–359.
5. James B.T. The effects of questionnaire mode on response in a federal employee survey: mail versus electronic mail. *Treat special contributed paper at the American statistical association annual conference*. 1997.
6. Dhaene G, Jochmans K.. Split-panel jackknife estimation of fixed-effect models, working paper. *Sciences Po departement of economics*. 2014.
7. Veiga A., Peter W.F.S., James J.B.. The use of sample weights in multivariate multilevel models with an application to income data collected by using a rotating panel survey. *Journal of the royal statistical society*. 2014; 63(1):65–84.
8. Patterson H.D.. Sampling on successive occasions with partial replacement of units. *Journal of the royal statistical society*. 1950; 12(1):241–255.
9. Eckler A.R.. Rotation sampling. *Annals of mathematical statistics*. 1955; 26(2):664–685.
10. Cochran W. *Sampling techniques*. New York: Wiley. 1977.
11. Nijman T.H.E., Verbeek M.. Estimation of time dependent parameters in linear models using cross-sections, panels or both [M]. *Journal of Economics* 1990, 46(1): 333–346.
12. Van Laarhoven P. J. M., Aarts E. H. L. *Simulated annealing: theory and applications*. Netherlands: Kluwer. 1987.
13. Campbell J.. Photosynthetic control of atmospheric carbonyl sulfide during the growing season. *Science*. 2008; 322(1):1085–1088.
14. Bailis R., Ezzati M., Kammen D.. Mortality and greenhouse gas impacts of biomass and petroleum energy futures in Africa. *Science*. 2005; 308(2):98–103.
15. Murphy J. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*. 2004; 430(1):768–772.
16. Saltelli A., Chan K., Scott E.M.. *Sensitivity analysis*. New York. 2000.
17. Saltelli A., Annoni P.. How to avoid a perfunctory sensitivity analysis. *Environmental modeling and software*. 2010; 25(1):1508–1517.
18. Nijman T.H.E., Verbeek M., Soest A.V.. The efficiency of rotating-panel designs in an analysis of variance model. *Journal of econometrics*. 1991; 49(1):373–399.
19. Duncan G.J., Juster F.T., Morgan J.N.. The role of panel studies in research on economic behavior. *Transportation research*. 1987; 21(2):249–263.