G.Ü. Fen Bilimleri Dergisi
18(4): 591-601(2005)
ISSN 1303-9709
G.U. Journal of Science
18(4): 591-601 (2005)

CALIBRATION ESTIMATOR

Aylin ALKAYA
Gazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, 06500, Ankara, TÜRKİYE

Alptekin ESİN*
Gazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, 06500, Ankara, TÜRKİYE,
e-mail: alpesin@gazi.edu.tr

ABSTRACT

A method which is similar to regression estimation method is the calibration estimator. The calibration estimator uses auxiliary variable(s) information to produce efficient estimates. Calibration requires that we know population totals for one or more auxiliary variable (x variables). The efficiency of the calibration estimator depends on how well the auxiliary variables explain the variability of y, the variable of interest. To improve the quality of estimates in sample surveys some kind of weighting is often carried out. This article reviews the calibration estimator which is one of the weighting method and attempts to show calculation of calibration weights with a hypothetical data.

***Key Words*:** Calibration, Weighting, auxiliary information, regression estimator

AYARLAMA TAHMİN EDİCİSİ

ÖZET

Regresyon tahmin yöntemine benzeyen bir diğer tahmin yöntemi de ayarlama tahmin edicisi yöntemidir. Ayarlama tahmin edicisi yardımcı değişken(ler) bilgisini tahminlerin etkinliğini arttırmak için kullanır. Ayarlama tahmin edicisi, bir veya birden çok yardımcı değişken (x değişkenleri) için yığın toplamlarını bilmeyi gerektirir. Bu tahmin edicinin etkinliği yardımcı değişken veya değişkenlerin ilgilenilen değişken y'yi ne kadar iyi açıklayabildiğine bağlıdır. Örnekleme araştırmalarında tahminlerin kalitesini arttırmak için bazı ağırlıklandırmalar sıklıkla kullanılmaktadır. Bu çalışmada ağırlıklandırma yöntemlerinden biri olan ayarlama tahmin edicisi incelenmeye çalışılmış, bir hipotetik veri ile ayarlanmış ağırlıkların nasıl hesaplandığı gösterilmeye çalışılmıştır.
Anahtar Kelimeler: Ayarlama, Ağırlıklandırma, Yardımcı bilgi, Regresyon tahmin edicisi

1. GİRİŞ

Örneklemede yüksek duyarlılığa sahip tahmin ediciler yardımcı değişken(ler) x_i (x_i'ler) hakkında güçlü bilgilere gereksinim duyar. N sayıda birimden oluşan bir yığından tesadüfi olarak seçilen örnek birimlerinin ilişkili x_i (x_i'ler) ve y_i ile gösterilen iki ölçümü elde edilebiliyorsa, bu değişkenlerden biri yardımıyla (yardımcı değişken(ler) ile) diğerinin tahmin edilmesi, tahminin duyarlılığını artırabilmek için başvurulan bir yöntemdir.

Ayarlama (**calibration**) tahmin edicileri Deville ve Särndal (1) tarafından anket verisinde yardımcı bilginin (x_i'lerin) kullanımı kapsamında ele alınmıştır. Ağırlıklandırma yöntemlerinden biri olan ayarlama, örnekleme tasarım ağırlıklarının araştırma verisine uygulanması durumunda ilgilenilen farklı araştırma değişkenleri (y_i'ler) için model-yansız tahmin ediciler sağlar (2).

Örneğimizde yardımcı değişken veya değişkenlere

1. INTRODUCTION

In sampling, high sensitive estimators need strong informations about auxiliary variable(s) x_i (more than one x_i). If the sample units' which are chosen randomly among the N units of population and two related measure x_i (more then one x_i) and y_i of one of these variables can be obtained, estimating the other one by the aid (by auxiliary variable(s)) is a method of increasing the sensitivity of estimation.

The calibration estimators were introduced by Deville and Särndal (1) in the scope of using auxiliary information (x_i) from research data. The calibration estimator which is one of the weighting methods, gives model-unbiased estimators for discrete research variables of interest (y_i) when sampling design weights applied to the research data (2).

If an adequate population information about auxiliary ilişkin yığın hakkında yeterli bilgi mevcut ise, gözlemler üzerinde ağırlıkların uygulanması parametre tahminlerinin duyarlılığı arttırılabilir. Tahminler ağırlıklandırılmış gözlemler üzerinden yapılır. Platek ve Gray (1980) ve Lindström et al. (1979) ağırlıklandırmayı yanın düzeltilmesinde kullanılacak önemli bir yöntem olarak tanımlamıştır (3). Varyansdaki azalma yardımcı değişken bilgisinin ilgilenilen değişken ile ilgili olması durumuna bağlıdır, eğer yardımcı değişkenler ile ilgilenilen değişken arasında ilişki var ise, tahminlerin duyarlılığını yardımcı değişken bilgisini kullanarak arttırılabilir. Tahmin edicinin kalitesi ilgilenilen değişken ile yardımcı değişken arasındaki ilişkinin gücü ile saptanır.

Ayarlama tahmin edicisinde öncelikle ilgilenilen değişken y_i ile ilişkili olduğu belirlenen yardımcı değişken veya değişkenlerin gözlenen değerlerine ağırlıklar verilir. Bu ağırlıklara d_i örnekleme tasarım ağırlıkları denir. Araştırmanın başlangıcında belirlenecek olan $d_i=1/\pi_i$ ağırlıkları gözlemlere verilecek önem düzeyini ifade eder. Bu ağırlıkların belirlenmesi araştırmadan araştırmaya değişir, ağırlıklar araştırmacının deneyim ve bilgisine dayalı olarak her zaman elde edilebilir veya bilinebilir (4).

Araştırmalarda ağırlıkların tespit edilmesi ilgilenilen değişken y_i'ye bağlı değildir, yalnızca yardımcı değişken(ler)e bağlıdır. Dolayısıyla, herhangi bir araştırmada ağırlıkları hesaplamada yalnızca yardımcı değişken bilgisine ihtiyaç vardır. Ağırlıklar örneği dengeler, bu nedenle yardımcı değişkenlerin örnek dağılımı yığın dağılımına uyar (3).

Tahmin etme, ağırlıkları araştırma verisiyle ilişkilendirilmeyle eşdeğerdir. Bu çalışmada verilen yardımcı değişken bilgisi ile ağırlıklandırma sistemi üzerinde odaklanmış olup, uzaklık ölçümü ve ayarlama denklemleri kümesi yardımıyla bir ağırlıklandırma sistemi uygulanmıştır. Ayarlama tahmin edicisi, orijinal örnekleme tasarım ağırlıkları ($d_i=1/\pi_i$) ile ayarlama denklemleri arasında verilen uzaklık ölçümünün mümkün olduğunca birbirine çok yakın olmasını sağlayacak şekilde ayarlanmış ağırlıkları kullanır. Her uzaklık ölçümüne karşılık gelen bir ayarlanmış ağırlık ve bir ayarlama tahmin edicisi vardır (1).

Örnekleme araştırmalarında parametre tahmini yardımcı bilgiye dayalı model-tabanlı yaklaşımlar olan regresyon ve oran tahminleri yoluyla yapılmaktadır. Literatürde model-tabanlı yaklaşımlar olarak sınıflandırılabilecek üç temel yöntem önerilmiştir: genel regresyon tahmin edicileri (GREG) (Cassel, Särndal ve Wretman 1976; Särndal 1980); ayarlama tahmin edicileri (Deville ve Särndal 1992); ve son zamanlarda önerilen ampirik olabilirlik yöntemleridir. Model-tabanlı ile tahmin edicilerin yaklaşık (asimptotik) olarak yansız tasarım olduğu ifade edilmektedir. Bu, çalışılan-modelin doğru veya yanlış olmasına bakılmaksızın ve eğer model doğru ise tahmin edicinin önemli düzeyde etkin olması demektir (5).

Çalışmamız kapsamında ayarlama tahmin edicisi incelenmeye çalışılmış, genelleştirilmiş regresyon tahmin edicisi için alternatif bir yöntem ayarlama yöntemine

variable or variables is available in our sample, assigning weights on the observations my increase the sensitivity of parameters' estimates. Estimates are obtained by summation of the weighted observations. Platek and Gray (1980) and Lindström et al. (1979) regard weighting as an important method in correcting the bias (3). The reduction in variance is relevant to the relation between auxiliary variables information and interest variable, if there is a relation between these variables then the sensitivity of estimations can be improved by the using auxiliary variables information. The quality of the estimator can determined by the power of the relation between interest variable and auxiliary variable.

At the first stage of calibration estimator, the weights will assigned to auxiliary variable or variables observed values which of them are correlated with interest variable y_i.

These weights are called as d_i sampling design weights. The $d_i=1/\pi_i$ weights which will be determined at the beginning of the research express the significance level of observations. The determination of these weights differs for each research, the weights can always obtained or known according to the researcher's experience and knowledge (4).

In researches determination of the weights aren't conditional upon to interest variable y_i, they are only conditional upon to auxiliary variable(s). In consequence, calculating the weights at any research we need only auxiliary variable information. The weights balance the sample, therefore the sampling distribution of auxiliary variables fits to population distribution (3).

Estimating is equivalent to associating weights with research data. In this article we focus on given auxiliary variable information and a weighting system has been applied by the aid of the distance measure and calibration equations set. The calibration estimator uses the calibration weights by providing the distance measure between original sampling design weights ($d_i=1/\pi_i$) and calibration equations as closer as possible. There is an equivalent calibrated weight and a calibration estimator for all distance measure (1).

The parameter estimation which is leaning against to auxiliary information in sampling researches is constructed by regression and ratio estimations which are defined as model-based approaches. There have been three main methods proposed in the literature which can be categorized as model-based approaches: general regression estimators (GREG) (Cassel, Särndal and Wretman 1976; Särndal 1980); calibration estimators (Deville and Särndal 1992); and more recently empirical likelihood methods. By model-based we mean that estimators are approximately (asymptotically) design unbiased. This means, irrespective of whether the working-model is correct or not and are particularly efficient if a working model is correct (5).

In the scope of our study we manage to present calibration estimator, an alternative method for generalized regression estimators is given according to the dayalı verilmiştir.

2. AYARLAMA TAHMİN EDİCİSİ

Ayarlama tahmin edicileri Deville ve Särndal (1) tarafından anket verisinde yardımcı bilgi kullanımı kapsamında verilmiştir. Yardımcı değişken x'in sonlu yığın toplamı (veya ortalaması) bilgisini kullanarak tahminler üzerinde avantaj sağlayan bir başka yöntem de ayarlama tahminidir (6).

Bir örneğe ilişkin verilerden ilgilenilen y değişkeninin toplamı $Y=\sum_{i=1}^{N} y_i$ tahmin edilmek istensin. $U=\{1,2,....,N\}$ sonlu yığın birimlerini, s yığından tesadüfi olarak seçilen bir örneği $(s \subseteq U)$ ve $P(s)$ herhangi bir s biriminin örneğe seçilmesi olasılığını gösteriyor olsun. $\pi_i = P(i \in s)$ i biriminin s örneğinde yer alması olasılığı olarak tanımlanmış olsun. $i \in s$ birimleri için, (y_i, x_i) gözlemlendiği ve x'in yığın toplamı $X=\sum_U x_i$ 'in bilindiği varsayılsın. $s=\{1,2,....,n\}$ örnek birimleri için $\pi_i = P(i \in s)$ içerme olasılıkları biliniyor ise Y yığın toplamı tahmini verilen bir uzaklık ölçümü ve ağırlıklandırma yapılarak ayarlama tahmin edicisi yöntemiyle elde edilebilir. Burada örnekleme tasarım ağırlıkları $d_i=1/\pi_i$, $d_i>1$ dir. Y'nin ayarlama tahmin edicisi;

$$\hat{Y}_C = \sum_{i \in s} w_i y_i \qquad [1]$$

burada bilinmeyen w_i: *ayarlama ağırlığı* şeklinde tanımlanmıştır ($i=1,...,n$). [1] nolu ayarlama denkleminde w_i yardımcı değişkene bağlı yapılan ağırlıklandırmayı ifade eder. Yardımcı değişkeninin bilinen yığın toplamı X, ayarlanmış ağırlıklara bağlı olarak aşağıdaki gibi tanımlanmıştır.

$$X = \sum_{i \in s} w_i x_i = \sum_{i=1}^{N} x_i \qquad [2]$$

Burada $\sum_{i \in s} w_i = N$ dir. Bir örnek araştırmasında her bir örnek birimi için, k ($j=1,..,k$) sayıda değişken bilgisi mevcutsa, x_i i inci birime ait bilgi $\underline{x}' = [x_{i1} ... x_{ik}]'$ vektörüyle gösterilebilir. x_{ik}: i ıncı birimin k değişken değerini ifade eder. Burada araştırmada yer alan tüm değişkenler değil yalnızca y değişkeniyle ilişkili olan k sayıdaki x yardımcı değişkenleri ele alınır.

Yukarıda da ifade edildiği gibi örnekleme tasarım ağırlıkları araştırmadan araştırmaya değişir, araştırmacının deneyim ve bilgisine dayalıdır. Ayrıca örnekleme tasarım ağırlıklarının veri toplamadan sorumlu kişi tarafından belirlendiği de varsayılabilir. d_i ağırlıkları yığından tesadüfi olarak seçilen örnekte yardımcı değişkenin yığın toplamlarını tahmin etmede kullanılabilir.

Yardımcı değişken yığın toplamı tahmini örnekleme tasarım ağırlıklarına dayalı aşağıdaki gibidir.

$$\hat{X} = \sum_{i=1}^{n} d_i x_i \qquad [3]$$

calibration method.

2. CALIBRATION ESTIMATOR

The notion of calibration estimators was introduced by Deville and Särndal (1) in the context of using auxiliary information from survey data. Another method which provides advantages on estimates by using auxiliary variable x's finite population total (or mean) information is calibration estimate (6).

Let the population total $Y=\sum_{i=1}^{N} y_i$ of interest variable y is wanted to estimate. Suppose $U=\{1,2,....,N\}$ is the set of labels for the finite population, let s be the sample which has selected randomly from the population $(s \subseteq U)$ and let $P(s)$ be the probability of selecting any s unit to the sample. $\pi_i = P(i \in s)$ is defined as the inclusion probability of the ith unit in s sample. Let (y_i, x_i) values are observed for $i \in s$ units the population total of x, $X=\sum_U x_i$ has known. If $\pi_i = P(i \in s)$ inclusion probabilities for $s=\{1,2,....,n\}$ sampled units are known, then a given distance measure and assigning weights procedure the population total of Y's estimation can derived by calibration estimator method. Here $d_i=1/\pi_i$, $d_i>1$ are sampling design weights. Y's calibration estimator;

$$\hat{Y}_C = \sum_{i \in s} w_i y_i \qquad [1]$$

where the unknown w_i is defined as a *calibrated weight* (i=1,...,n). the w_i calibration system in [1] represents the weighting procedure which was constructed by according to the auxiliary variable. The known population total of auxiliary variable X, which has been determined according to *calibrated weights* is given as:

$$X = \sum_{i \in s} w_i x_i = \sum_{i=1}^{N} x_i \qquad [2]$$

where $\sum_{i \in s} w_i = N$. For each sample unit in a sample research, if the k ($j=1,..,k$) variables information is available, ith unit information of x_i can be illustrated by vector $\underline{x}' = [x_{i1} ... x_{ik}]'$. x_{ik} denotes kth variable value for ith unit. Here, we undertake only the x auxiliary variables in number of k which are related to the y interest variable not all variables which takes place in research.

As we mentioned before, the sampling design weights changes from one research to another, they depend on researcher's experience and knowledge. Moreover it is assumed that each sampling design weights are provided by responsible for collecting the data. Such d_i weights might be used to estimate the population total of auxiliary variable from the samplewhich has been selected randomly from the population.

The estimation of the population total of auxiliary variable according to the sampling design weights is as follows:

d_i: *örnekleme tasarım ağırlıkları, i=1,...,n* dır. w_i ağırlıklarının hesaplanmasında, x yardımcı değişkenleri için X yığın toplamları bilinir. Hesaplamada yapılması gereken yeni w_i ağırlıklarını mümkün olduğunca d_i tasarım ağırlıklarına yakın olmasını sağlamaktır. Bu nedenle, [2] ve [3] nolu denklemler ile verilen iki ağırlıklandırma kümesi arasındaki yakınlığı saptamaya yarayacak bir kriter belirlemek gereklidir. Bu kriter w_i ile d_i arasındaki uzaklıktır. Genel olarak, w_i ve d_i arasındaki uzaklık $G(w_i,d_i)$ ile gösterilir. Örnekleme tasarım ağırlıkları ile ayarlanmış ağırlıkları arasındaki toplam uzaklık;

$$U=\sum_{i}^{n} G(w_i,d_i) \quad [4]$$

şeklinde ifade edilebilir (7)[1]. Dolayısıyla sorun, [3] nolu eşitliğe dayalı olarak toplam uzaklığı minimum yapmaktır. Bunun için Lagrange çarpanı kullanılır.

$$L=\sum_{i}^{n} G(w_i,d_i)+\sum_{j=1}^{k}\lambda_j\left[X_j-\sum_{i=1}^{n} w_i x_{ji}\right] \quad [5]$$

λ_j *: Lagrange çarpanları, j=1,..k*

Amaç ayarlanmış ağırlıkları hesaplamak olduğundan, L denklemini minimum yapacak şekilde w_i değerlerini, $\sum_{i}^{n} G(w_i,d_i)$ için ki-kare uzaklık ölçümünü yazarak elde edebiliriz.

Burada uzaklık ölçümü olarak araştırmacılar tarafından yaygın kullanımı olan ki-kare uzaklık ölçümü kullanılmıştır. Yeniden ağırlıklandırmada ki-kare uzaklık ölçümü yerine diğer uzaklık ölçümleri de kullanılabilir. Ancak Deville ve Särndal (1) diğer uzaklık ölçümlerinin ki-kare ölçümüyle hemen hemen aynı sonuçları verdiğini ve ki-kare uzaklık ölçümüne asimptotik olarak denk olduklarını göstermiştir. Alternatif uzaklık ölçümleri Deville ve Särndal (1)'ın çalışmasında detaylı bir şekilde ele alınmıştır. Ki-kare uzaklık ölçümü;

$$\sum_{i}^{n} G(w_i,d_i)=\frac{1}{2}\sum_{i=1}^{n}\frac{(w_i-d_i)^2}{d_i} \quad [6]$$

şeklindedir. Ki-kare uzaklık ölçümü [5] nolu eşitlikte verilen Lagrange çarpanında yerine yazılırsa aşağıdaki eşitliğe ulaşılır.

$$L=\frac{1}{2}\sum_{i=1}^{n}\frac{(w_i-d_i)^2}{d_i}+\sum_{j=1}^{k}\lambda_j\left[X_j-\sum_{i=1}^{n} w_i x_{ji}\right] \quad [7]$$

Burada λ_j : Lagrange çarpanları, $j=1,..k$, X_j : j. birimin bilinen X yığın toplamı vektörünü göstermektedir. Optimum w_i, [7] nolu eşitlikte L'nin w_i'ye göre türevinin alınıp sıfıra eşitlenmesiyle aşağıdaki gibi elde edilir.

$$\frac{\partial L}{\partial \mathrm{w_i}}=\left[\frac{w_i}{d_i}-1\right]-\sum_{j=1}^{k}\lambda_j x_{ji}=0 \quad [8]$$

$$\hat{X}=\sum_{i=1}^{n} d_i x_i \quad [3]$$

d_i: *sampling design weights, i=1,...,n.* In calculation of w_i weights we know the population total of auxiliary variable X. The thing to do in computing is to ensure the new weights w_i's as close as possible to the design weights d_i. Thus, a constraint which will be used to find the closeness between two weighting set given by [2] and [3] equations should be determined. This constraint is the distance between w_i and d_i. In general, the distance between w_i and d_i is denoted as $G(w_i,d_i)$. The total distance between sampling design and calibrated weights is as follows (7)[1]:

$$U=\sum_{i}^{n} G(w_i,d_i) \quad [4]$$

The problem is therefore to minimize the total distance [4] subject to the equation [3]. For this problem the Lagrange multiplier is used.

$$L=\sum_{i}^{n} G(w_i,d_i)+\sum_{j=1}^{k}\lambda_j\left[X_j-\sum_{i=1}^{n} w_i x_{ji}\right] \quad [5]$$

λ_j : *Lagrange* multipliers, $j=1,..k$

Our aim is to calculate calibrated weights, so obtaining the values of w_i by minimizing the L equation, we use chi-squared distance measure as $\sum_{i}^{n} G(w_i,d_i)$. Here for distance measure, the chi-squared distance measure which is the most commonly used by the researchers was used. Alternative distance measures can also be considered for reweighting instead of chi-squared distance measure. But it has shown by Deville and Särndal (1) that the alternative distance measures gives almost the same results as chi-squared measure and are asymptotically equivalent to the one by using the chi-squared distance measure. The detailed information about alternative distance measures were given in Deville and Särndal (1)'s study. The chi-squared distance measure is:

$$\sum_{i}^{n} G(w_i,d_i)=\frac{1}{2}\sum_{i=1}^{n}\frac{(w_i-d_i)^2}{d_i} \quad [6]$$

If we replace chi-squared distance measure in the Lagrange multiplier equation [5] we will get the equation below.

$$L=\frac{1}{2}\sum_{i=1}^{n}\frac{(w_i-d_i)^2}{d_i}+\sum_{j=1}^{k}\lambda_j\left[X_j-\sum_{i=1}^{n} w_i x_{ji}\right] \quad [7]$$

where the λ_j for $j=1,..k$ are Lagrange multipliers and X_j represents the jth element of the vector of known population total. Differentiation of (7) by w_i, and then equal to zero gives the optimum value of w_i as below.

$$\frac{\partial L}{\partial \mathrm{w_i}}=\left[\frac{w_i}{d_i}-1\right]-\sum_{j=1}^{k}\lambda_j x_{ji}=0 \quad [8]$$

1 Bazı yazarlar, Falson ve Singh (2000) gibi minimize edilecek uzaklığı $\sum_i^n d_i G(w_i,d_i)$ olarak yazmaktadır, ancak bu çalışmada Deville ve Särndal (1992) izlenmektedir.

1 Some authors, such as Falson and Singh (2000)write the distance to be minimized as $\sum_i^n d_i G(w_i,d_i)$, but the present paper follows Deville and Särndal (1992).

$\frac{\partial L}{\partial \mathrm{w_i}}$ eşitliğinde $\sum_{j=1}^{k} \lambda_j x_{ji}$ yerine $x_i'\lambda$ yazılabilir.

Buradan, [8] nolu eşitlikte w_i yalnız bırakılırsa,

$$w_i = d_i(1 + x_i'\lambda),\ i=1,\dots,n \qquad [9]$$

sonucuna ulaşılmış olur. Burada λ değerine ihtiyaç vardır. [9] nolu eşitliğinin her iki tarafı x_i ile çarpıldığında,

$$w_i x_i - d_i x_i = d_i x_i x_i' \lambda \qquad [10]$$

eşitliğine ulaşılır. [10] nolu eşitliği tüm n'ler üzerinden toplarsak aşağıdaki eşitlik elde edilmiş olur.

$$X - \hat{X} = \left[\sum_{i=1}^{n} d_i x_i x_i'\right]\lambda \qquad [11]$$

λ'nın hesaplanabilmesi için λ'nın yalnız bırakılması gereklidir. Bunun için [11] nolu eşitliğin her iki tarafı soldan $\left[\sum_{i=1}^{n} d_i x_i x_i'\right]^{-1}$ ters matrisi ile çarpıldığında λ lagrange çarpanları vektörüne ulaşılmış olur.

$$\lambda = \left[\sum_{i=1}^{n} d_i x_i x_i'\right]^{-1} \left(X - \hat{X}\right) \qquad [12]$$

Elde edilen λ değerleri [9] nolu eşitlikte yerine yazılacak olursa yeni ağırlıklar (ayarlanmış ağırlıkları) elde edilmiş olacaktır. [9] nolu denklemi $w_i = d_i(1 + \lambda' x_i)$ şeklinde yazmak da mümkündür, burada $\lambda' = (X - \hat{X})' T^{-1}$ olup T simetrik matrisi $\sum_{i=1}^{n} d_i x_i x_i'$ eşittir.

Ayarlanmış ağırlıkları $i=1,\dots n$ için hesaplandığında, Y yığın toplamının tahmini [1] nolu ayarlama tahmin edicisiyle bulunabilir.

$$\hat{Y}_C = \sum\nolimits_{i \in s} w_i y_i \qquad [1]$$

$\hat{Y}_C$ ayarlama tahmin edicisinde [9] nolu denklem ile verilen w_i ayarlanmış ağırlıkları denklemi yerine yazılacak olursa,

$$\hat{Y}_C = \sum_{i=1}^{n} d_i y_i + (X - \hat{X})' B, \qquad [13]$$

sonucuna ulaşılır, $B = T^{-1} \sum_{i=1}^{n} d_i x_i y_i$. Böylece, ayarlama tahmin edicisi ile genelleştirilmiş regresyon (GREG) tahmin edicisi arasında bir bağlantı kurulmuş olur.

Ayarlanmış ağırlıklarının hesaplamaları basit bir örnekle verilmeye çalışılmıştır.

ÖRNEK. Araştırma değişkeni ile ilişkili ve yığın toplamı değerleri bilinen üç yardımcı değişken olduğu varsayılsın. $n=30$ örnek birimi için hipotetik veri Çizelge 1'deki gibi

In equation $\frac{\partial L}{\partial \mathrm{w_i}}$ instead of $\sum_{j=1}^{k} \lambda_j x_{ji}$, we can substitute $x_i'\lambda$. Hence, from equation we will obtain w_i:

$$w_i = d_i(1 + x_i'\lambda),\ i=1,\dots,n \qquad [9]$$

Here the value of λ is required. Multiplication of two side of the equation [9] by x_i gives:

$$w_i x_i - d_i x_i = d_i x_i x_i' \lambda \qquad [10]$$

Summing [10] over all n, the following equation will be obtained.

$$X - \hat{X} = \left[\sum_{i=1}^{n} d_i x_i x_i'\right]\lambda \qquad [11]$$

To obtain λ it will have to be let alone. Therefore, when two side of the equation [11] from the left hand side multiplied by inverse matrix $\left[\sum_{i=1}^{n} d_i x_i x_i'\right]^{-1}$ the vector of Lagrange multiplier is given by:

$$\lambda = \left[\sum_{i=1}^{n} d_i x_i x_i'\right]^{-1} \left(X - \hat{X}\right) \qquad [12]$$

The resulting values of λ are substituted into [9] to obtain the new weights (calibrated weights). It can be possible to write equation [9] as $w_i = d_i(1 + \lambda' x_i)$, here $\lambda' = (X - \hat{X})' T^{-1}$ and the symmetric matrix of T is equal to $\sum_{i=1}^{n} d_i x_i x_i'$.

When the calibrated weights computed for $i=1,\dots n$, the estimation of the population total of Y can be obtained by calibration estimation which was given by equation [1].

$$\hat{Y}_C = \sum\nolimits_{i \in s} w_i y_i \qquad [1]$$

Substituting the calibrated weights equation w_i which was given by equation [9] in calibration estimation $\hat{Y}_C$, gives:

$$\hat{Y}_C = \sum_{i=1}^{n} d_i y_i + (X - \hat{X})' B, \qquad [13]$$

where $B = T^{-1} \sum_{i=1}^{n} d_i x_i y_i$. Thus, a correlation has been established between calibration estimation and general regression estimator (GREG).

The calculations of the calibrated weights was managed to illustrated by a simple example.

EXAMPLE. Suppose there are three auxiliary variables correlated with the research variable and for which population total values are available. Let the hypothetical data for a sample of $n=30$ sample units are shown in Table 1. Suppose the auxiliary variables were defined as below.

olsun. Yardımcı değişkenler aşağıdaki gibi tanımlanmış olsun. *X_1: Kişi Kadın ise 1, Erkek ise 0; X_2: Kişi genç ise 1, değil ise 0 değerini alıyor; X_3: Çalışma statüsünü gösteriyorken kişi işçi ise 1, memur ise 0* değerini aldığı kabul edilsin. Yardımcı değişkenler için bilinen yığın toplamları ***X=[X_1 X_2 X_3]'= [30,60,70]'*** şeklinde varsayılmış olsun. Yardımcı değişkenlerin örnek toplamaları, Çizelge 1'de 5. kolonda verilen d_i örnek tasarım ağırlıkları ile $\hat{X}=\sum_{i=1}^{n} d_i x_i$ formülünden yararlanılarak, $\hat{X}$ ***=[35,68,67]'*** şeklinde hesaplanmıştır.

Yukarıda verilen yardımcı değişkenlerin yığın toplamı değerleri örnekteki değerler ile kıyaslandığında, yığında örnektekine kıyasla daha az sayıda kadın, daha az sayıda genç ve daha çok sayıda işçi olduğu görülür. Örnekleme tasarım ağırlıklarının aslında daha çok sayıda erkek, daha az sayıda genç ve daha çok sayıda işçiyi yansıtacak şekilde oluşturulması istenilirdi. Yardımcı değişkenlerin yığın toplamı bilgisini daha iyi yansıtmak amacıyla yeniden ağırlıklandırma yapılmasını öneren ayarlama tahmin edicisiyle, w_i ayarlanmış ağırlıkları hesaplanır.

Ayarlama ağırlıkları hesaplanmıştır ve Çizelge 1'in son kolonunda verilmiştir. $\sum_{i=1}^{n} d_i x_i x_i'$ simetrik matrisi ve bu matrisin tersi Çizelge 2'de verilmiştir.

X_1:1 for those who are Female, 0 who are Male; X_2: 1 for those who are young, 0 otherwise; Variable 3 refers to personal statute so that X_3:1 for those who are blue-collar worker, 0 who are civil servant. Let *t*he known population totals for auxiliary variables were assumed as ***X=[X_1 X_2 X_3]'= [30,60,70]'***. The sample total of auxiliary variables were calculated as $\hat{X}$ ***=[35,68,67]',*** by the aid of sampling design weights d_i which was shown in the 5 column of Table 1 and formula $\hat{X}=\sum_{i=1}^{n} d_i x_i$.

When the population total values of auxiliary variables which were given as above were compared with the values in sample, it has seen that there are fewer females, fewer younger units but more blue-collar worker in population than in sample. In fact the sampling design weights were desired to reflect more male, few younger unit and more blue-collar worker. By the calibration estimation which propose reweighting to reflect population total values of auxiliary variables better, the calibrated weights w_i will be calculated.

The calibrated weights were calculated and shown in the last column of Table 1. $\sum_{i=1}^{n} d_i x_i x_i'$ is a symmetric matrix and its inverse are given in Table 2.

Table 1. Sample values of auxiliary variables and calibrated weights
Çizelge 1. Yardımcı değişkenlerin örnek değerleri ve ayarlanmış ağırlıklar

i	X_1	X_2	X_3	d_i	w_i
1	1	1	1	4	4,958
2	0	1	1	5	5,085
3	1	0	0	2	2,445
4	1	0	0	2	2,445
5	0	1	0	4	3,369
6	0	0	1	4	4,699
7	0	0	0	3	3,000
8	0	0	0	3	3,000
9	0	1	1	5	5,085
10	1	1	0	3	3,195
11	1	1	1	4	4,958
12	1	0	1	3	4,192
13	1	0	1	3	4,192
14	0	1	1	5	5,085
15	0	0	1	4	4,699

i	X_1	X_2	X_3	d_i	w_i
16	1	0	0	2	2,445
17	0	1	1	5	5,085
18	0	1	0	4	3,369
19	1	0	0	2	2,445
20	0	0	0	3	3,000
21	0	0	1	4	4,699
22	0	0	1	4	4,699
23	1	1	0	3	3,195
24	0	1	1	5	5,085
25	0	1	0	4	3,369
26	1	1	1	4	4,958
27	0	1	1	5	5,085
28	1	0	1	3	4,192
29	0	1	0	4	3,369
30	0	1	0	4	3,369

Table 2. Matrix $\sum_{i=1}^{n} d_i x_i x_i'$ and its inverse

Çizelge 2. Matris $\sum_{i=1}^{n} d_i x_i x_i'$ ve tersi

	$\sum_{i=1}^{n} d_i x_i x_i'$	
35	*18*	*21*
18	*68*	*33*
21	*33*	*67*

	$\left[\sum_{i=1}^{n} d_i x_i x_i'\right]^{-1}$	
0,0366	*-0,0054*	*-0,0088*
-0,0054	*0,0200*	*-0,0082*
-0,0088	*-0,0082*	*0,0217*

w_i ayarlanmış ağırlıkları, $w_i = d_i(1 + x_i'\lambda)$ denkleminde $\lambda = \left[\sum_{i=1}^{n} d_i x_i x_i'\right]^{-1} \left(X - \hat{X}\right)$ yazılmasıyla Çizelge 1 ve Çizelge 2'deki değerlerden hesaplanır.

The calibrated weights w_i will be obtained by the values in Table 1 and Table 2, writing $\lambda = \left[\sum_{i=1}^{n} d_i x_i x_i'\right]^{-1} \left(X - \hat{X}\right)$ value in $w_i = d_i(1 + x_i'\lambda)$ equation.

$$\left(X - \hat{X}\right) = \begin{bmatrix} -5 \\ -8 \\ 3 \end{bmatrix}, \quad \lambda = \left(\begin{array}{|c|c|c|} \hline 0,0366 & -0,0054 & -0,0088 \\ \hline -0,0054 & 0,0200 & -0,0082 \\ \hline -0,0088 & -0,0082 & 0,0217 \\ \hline \end{array} \right) \begin{bmatrix} -5 \\ -8 \\ 3 \end{bmatrix},$$

$$\lambda = \begin{bmatrix} 0,2226 \\ -0,1576 \\ 0,1747 \end{bmatrix}$$

Ayarlanmış ağırlıkları $w_i = d_i\ (1 + x_i'.\lambda)$ formülünden $i = 1, \ldots, 30$'a kadar tüm birimler için elde edilir. Örneğin Çizelge 1'deki w_1 ağırlığı şöyle bulunmuştur.

$$w_1 = 4\{\ 1 + (1\ \ 1\ \ 1) \begin{bmatrix} 0,2226 \\ -0,1576 \\ 0,1747 \end{bmatrix} \} = 4,958$$

The calibrated weights, for all units $i = 1, \ldots, 30$, obtained by $w_i = d_i\ (1 + x_i'.\lambda)$ formula. For example the w_1 weight in Table 1 was obtained as:

$$w_1 = 4\{\ 1 + (1\ \ 1\ \ 1) \begin{bmatrix} 0,2226 \\ -0,1576 \\ 0,1747 \end{bmatrix} \} = 4,958$$

Ağırlıklarda yapılan gerekli düzeltmelerin beklenen değerlerle tutarlı olduğu gözlenmektedir. Ayarlama için gerekli bilgiler mevcutsa kişilerin özelliklerini (karakteristiklerini) de yansıtan ağırlıklandırmanın yapılması yürütülecek tahminlerin duyarlılığını arttıracağından, tahmin için ayarlama tahmin edicisi kullanılması yerinde bir karar olacaktır. Bu örneğimiz için Çizelge 1'de yer alan *5, 18, 25, 29, 30* numaralı kişilerine ilişkin ağılıkları inceleyelim. Bu kişiler için öngörülen ayarlanmış ağırlıkları ile kişilerin karakteristik özellikleri de göz önüne alındığında, ayarlanmış ağırlıkları d_i örnek ağırlıklarına kıyasla göreli olarak düşük bulunmuştur (*4*'ten *3,369*'a düşmüştür). Örnektekine kıyasla yığında çalışan işçi sayısı yüksek olmasına karşın kadın ve genç sayısının daha az olması nedeniyle *5,18,25,29,30* numaralı kişilere verilecek ağırlıkta daha düşük olmalıdır, bu ise ayarlanmış ağırlıklarla sağlanmıştır.

The required adjustments made on the weights are observed to be consistent with the expected values. If the required information for calibration is available, constructing the weights as reflecting the units characteristics will increase sensitivity of estimations, so using calibration estimation for estimating will be an appropriate decision. For our example, lets examine the weights for number of *5, 18, 25, 29, 30* units in Table 1. When the calibrated weights which are prescribed for the units and the characteristics of the units are taken into consideration, the calibrated weights found relatively low in comparison with the sampling design weights d_i (fall from *4* to *3,369*). Although the number of blue-collar worker in population is higher than in sample, the weights given to the number of *5,18,25,29,30* should be lower because of the few number of female and young units, this situation was provided by calibrated weights.

3. NEDEN AYARLANMIŞ AĞIRLIKLAR?

Bir çok araştırma istatistikçisi ayarlanmış ağırlık yönteminin aşağıda yer alan özellikleri taşıdığına katılmaktadır (8):

1. *Tutarlılık.* Ayarlama tahmin edicisi tutarlı bir tahmin edicidir. Bu yöntemde, [2] nolu denklemi sağlayan bir ağırlıklandırma sistemi uygulanır. Böylece her bir yardımcı değişkeninin bilinen yığın toplamı değerlerinin en iyi şekilde temsil edilmesi sağlanmış olur.
2. *Temel ağırlıklara çok yakın olma.* Temel (orijinal) örnekleme tasarım ağırlıkları $d_i=1/\pi_i$ 'nin en (cazip) önemli özelliği model yansız tahminler vermesidir. Dolayısıyla, bu ağırlıklardan olabilecek herhangi bir sapma az olmalıdır ki model yansızlık en azından yaklaşık veya asimptotik olarak söz konusu olsun.
3. *Yardımcı değişken toplamları üzerinde kontrol sağlama.* Ayarlamada ilgilenilen değişken ile ilişkili olduğu bilinen ne kadar çok yardımcı değişken bilgisi kullanılmışsa o kadar "iyi" düzeyde ağırlıklandırma sistemi yapılmış demektir. Ayarlama tahmin edicilerin duyarlılığı y değişkeni ile ilişkili olan çok sayıda (otokorelasyon sorunu olmayan ve yığın toplamı bilinen) yardımcı değişken kullanımıyla artış gösterecektir.

4. GENEL REGRESYON TAHMİN EDİCİLERİNİN AYARLAMA İLE ELDE EDİLMESİ

Ayarlama yaklaşımına olan ilgi Deville and Särndal (1992)'ın ayarlamanın genel regresyon tahmin edicisine olan asimptotik denkliğini göstermesiyle artmıştır. Genel regresyon tahmin edicileri (GREG) çok değişkenli yardımcı bilgi ile tasarlanmıştır. Burada, GREG'in ağırlıklar üzerinde odaklanılarak farklı bir yoldan elde edilebileceği gösterilmiştir.

i inci gözlemin örnekleme tasarım ağırlıkları d_i dir. GREG de ki ağırlıklar, verilen bir uzaklık ölçümüne dayalı bir şekilde, $d_i=1/\pi_i$ 'ye olabildiğince yakın tutulmaktadır. Bu durum ağırlıklandırılmış yardımcı değişken değerlerinin örnek toplamının, X yığın toplamını bu yardımcı değişkenlere eşit yapması gerektiğini ifade eder. Bu, ayarlanmış ağırlıklar her bir yardımcı değişkene uygulandığında mükemmel tahminler vermelidir demektir. Dolayısıyla, bu da tutarlılığı sağlar, çünkü yardımcı değişken ile ilgilenilen değişken arasında güçlü bir ilişkinin olması yardımcı değişken için önemli bir rol oynayan ağırlıkların ilgilenilen değişken için de önemli olacağı anlamına gelir. GREG yardımcı bilgiyi etkin kullanır, bu nedenle tahminler kusursuzdur; buna karşın birey ağırlıkları her zaman uygun olamamaktadır (1).

Ayarlama tekniği, genel regresyon tahmin edicisi (GREG) için bir alternatif çıkarsama olup [14] nolu denklem ile verilmiştir. Y'nin ayarlama tahmin edicisi regresyon tahmin edicisine bir alternatif olacak şekilde aşağıdaki gibi yazılabilir.

3. WHY CALIBRATED WEIGHTS?

Many survey statisticians agree on the following objectives for a calibrated weight method (8):

1. *Consistency.* Calibration estimation is a consistent estimation. In this method, a weighting system that satisfies [2] is applied. Thereby, it reproduces the representation of the known population total value of auxiliary variable at optimum way.
2. *Closeness to the basic weights.* The most attractive property of the basic sampling design weights have is that they yield design unbiased estimates. Therefore, any departure from these weights should be small in order to preserve the design unbiasedness, at least approximately or asymptotically.
3. *Control on auxiliary variable totals.* The more auxiliary variables information we use in calibration, which are known to be related with interest variable, the "better" we expect the resulting weight system to be. The sensitivity of calibration estimations tends to increase as more auxiliary variables (known that have no autocorrelation problem and known population total) which are correlated with interest variable are used.

4. DERIVING THE GENERAL REGRESSION ESTIMATORS BY CALIBRATION

The interest for the calibration approach has grown since Deville and Särndal (1992) showed the asymptotic equivalence of calibration to the general regression estimator. The general regression estimators (GREG) were conceived with multivariate auxiliary information. Here, it has shown that the GREG can be derived by a different route by focusing on the weights.

The sampling design weights of ith observation is d_i. The weights implied by the GREG are as close as possible to the $d_i=1/\pi_i$ according to a given distance measure. This situation signifies that the sample total of weighted auxiliary variable values must be equal to the known population total X for that auxiliary variable. That is, the calibrated weights must give perfect estimates when applied to each auxiliary variable. So this provides the consistency, because a strong correlation between the auxiliary variable and the interest variable means that the weights that perform well for the auxiliary variable also should perform well for the interest variable. GREG uses the auxiliary information efficiently, so the estimates are perfect; despite this the individual weights are not always suitable. (1).

The calibration method is an alternative inference for general regression estimator (GREG) and was given by [14]. The calibration estimator of Y can be written as below as an alternative to the regression estimator:

$$\hat{Y}_{C\_reg} = \sum_{i=1}^{n} w_i y_i = \hat{Y}_{HT} + (X - \hat{X}_{HT})'\hat{B} \qquad [14]$$

Burada $\hat{X}_{HT} = \sum_{i \in s} d_i x_i$ x vektörü için Horwitz-Thompson tahmin edicisidir ve

$$\hat{B} = \left\{\sum_{i \in s} d_i q_i x_i x_i'\right\}^{-1} \sum_{i \in s} d_i q_i x_i y_i \qquad [15]$$

çoklu regresyon katsayısının ağırlıklandırılmış bir tahmin edicisidir.

Deville ve Särndal (1) $\hat{Y}_C$ ayarlama tahmin edicisinin $\hat{Y}_{C\_reg}$ regresyon tahmin edicisine asimptotik olarak denk olduğunu çalışmasında göstermiştir.

5. VARYANS VE VARYANS TAHMİNİ

Yukarıda da ifade edildiği üzere $\hat{Y}_C$ ayarlama tahmin edicisi $\hat{Y}_{C\_reg}$ regresyon tahmin edicisine asimptotik olarak denktir. Dolayısıyla $\hat{Y}_C$ 'nin asimptotik varyansı (*AV*) regresyon tahmin edicisinin varyansıyla aynı olup, şöyledir (10):

$$AV(\hat{Y}_C) = \sum_{i=1}^{N}\sum_{l=1}^{N} \Delta_{il}(d_i E_i)(d_l E_l) \qquad [16]$$

$\Delta_{il} = \pi_{il} - \pi_i \pi_l$, $d_i = 1/\pi_i$ ve $E_i = y_i - x_i'B$ olup, B [17] nolu eşitliği sağlar.

$$\left(\sum_{i=1}^{N} q_i x_i x_i'\right)B = \sum_{i=1}^{N} q_i x_i y_i \qquad [17]$$

B ağırlıklandırılmış en küçük kareler ifadesini minimize eder:

$$SS_U = \sum_{i=1}^{N} q_i (y_i - x_i'B)^2 = \sum_{i=1}^{N} q_i E_i^{\,2} \,. \qquad [18]$$

[16] nolu eşitlikte yer alan varyansı tahmin etmek için, E_i artıkları kullanılamaz, çünkü B bilinmemektedir. $\hat{B}_{ws}$ tahmin edicisi, [18] nolu eşitlik ile verilen bilinmeyen $q_i E_i^2$ değerinin yığın toplamı olan SS_U ile hesaplanır. Bu toplamın ayarlanmış ağırlıklar tahmin edicisi $SS_{sw} = \sum_{i=1}^{n} w_i q_i E_i^{\,2}$ dir ve $\left(\sum_{i=1}^{n} w_i q_i x_i x_i'\right)\hat{B}_{ws} = \sum_{i=1}^{n} w_i q_i x_i y_i$ örnek-tabanlı normal denklemini sağlayan $\hat{B}_{ws}$ tahmin edicisi ile minimize edilir. Örnek-tabanlı artıklar $e_i = y_i - x_i'\hat{B}_{ws}$ olarak hesaplanabilir. $\hat{Y}_C$ 'nin varyans tahmin edicisi aşağıdaki gibi yazılır.

$$\hat{Y}_{C\_reg} = \sum_{i=1}^{n} w_i y_i = \hat{Y}_{HT} + (X - \hat{X}_{HT})'\hat{B} \qquad [14]$$

where $\hat{X}_{HT} = \sum_{i \in s} d_i x_i$ is the Horwitz-Thompson estimator for the *x*-vector and

$$\hat{B} = \left\{\sum_{i \in s} d_i q_i x_i x_i'\right\}^{-1} \sum_{i \in s} d_i q_i x_i y_i \qquad [15]$$

is a weighted estimation of the multiple regression coefficients.

Deville and Särndal (1) state that the calibration estimator $\hat{Y}_C$ is asymptotically equivalent to the regression estimator $\hat{Y}_{C\_reg}$.

5. VARIANCE AND VARIANCE ESTIMATION

As mentioned above, the calibration estimator $\hat{Y}_C$ is asymptotically equivalent to the regression estimator $\hat{Y}_{C\_reg}$. Thus, the asymptotic variance (*AV*) of $\hat{Y}_C$ is the same as that of the regression estimator and it is (10):

$$AV(\hat{Y}_C) = \sum_{i=1}^{N}\sum_{l=1}^{N} \Delta_{il}(d_i E_i)(d_l E_l) \qquad [16]$$

where $\Delta_{il} = \pi_{il} - \pi_i \pi_l$, $d_i = 1/\pi_i$ and $E_i = y_i - x_i'B$, B satisfying the equation [17].

$$\left(\sum_{i=1}^{N} q_i x_i x_i'\right)B = \sum_{i=1}^{N} q_i x_i y_i \qquad [17]$$

B minimizes the weighted least square expression:

$$SS_U = \sum_{i=1}^{N} q_i (y_i - x_i'B)^2 = \sum_{i=1}^{N} q_i E_i^{\,2} \,. \qquad [18]$$

To estimate the variance [16], the residuals E_i cannot be used, because B is unknown. The estimator $\hat{B}_{ws}$ is calculated by SS_U given by [18] is the population total of unknown $q_i E_i^2$ values. The calibrated weights estimator of this total is $SS_{sw} = \sum_{i=1}^{n} w_i q_i E_i^{\,2}$ and which is minimized by estimator $\hat{B}_{ws}$ satisfying the sample-based normal equation $\left(\sum_{i=1}^{n} w_i q_i x_i x_i'\right)\hat{B}_{ws} = \sum_{i=1}^{n} w_i q_i x_i y_i$. The sample-based residuals can be calculated as $e_i = y_i - x_i'\hat{B}_{ws}$. The variance estimator of $\hat{Y}_C$ can be written as below:

$$\hat{V}(\hat{Y}_C) = \sum_{i=1}^{n}\sum_{l=1}^{n} (\Delta_{il}/\pi_{il})(w_i e_i)(w_l e_l) \qquad [19]$$

$$\hat{V}(\hat{Y}_C)=\sum_{i=1}^{n}\sum_{l=1}^{n}(\Delta_{il}/\pi_{il})(w_i e_i)(w_l e_l) \quad [19]$$

Ayarlanmış ağırlıkları $\hat{V}(\hat{Y}_C)$'de e_i artıklarını ağırlıklandırmak için kullanılmıştır. Tasarım-tabanlı bakış açısıyla ele alınacak olunursa, e_i artıklar standart tasarım ağırlıkları d_i'lerden basit bir şekilde ağırlıklandırılabilir ve tasarım-tutarlı bir varyans tahmin edicisi elde edilir. Model-tabanlı, tasarım-tabanlıda ki özellikleri taşımasına karşın, d_i'lerin üzerine w_i'lerin tercih edilmesinin bir sebebi vardır. Regresyon tahmin edicisinin [14] temelini oluşturan model ξ, $E_\xi(y_i)=\beta' x_i$, $V_\xi(y_i)=\sigma_i^2$ olduğunu belirtir. Buradan, [19] nolu eşitliğin yalnızca tasarım-tutarlı bir varyans tahmin edicisi değil, aynı zamanda model hata kareler ortalaması $E_\xi(\hat{Y}_c - Y)^2$ için adeta bir model-yansız tahmin edicidir. Basit tesadüfi örneklemede yerine koymadan örnekleme yapıldığında, eğer örnekleme kesri küçük ise [14] nolu eşitliğin model yanı ihmal edilebilmektedir.

The calibrated weights were used in $\hat{V}(\hat{Y}_C)$ to weight the residuals e_i. From the design-based point of view, the residuals e_i can be weighted easily by the standard design weights d_i and a design-consistent variance estimator can be obtained. Altough model-based has the same properties with design-based properties, there is reason for preferring the w_i over d_i. The model ξ that underlies the regression estimator [14] states that $E_\xi(y_i)=\beta' x_i$, $V_\xi(y_i)=\sigma_i^2$. Hence, the equation is [19] not only a design-consistent variance estimator but also a model-unbiased for the model mean square error, $E_\xi(\hat{Y}_c - Y)^2$ (1).

6. SONUÇ

Araştırmacılar yardımcı değişkenler için mükemmel tahminler sağlayan ağırlıkların, ilgilenilen değişken içinde iyi bir tahmin sağlayacağına inanmaktadırlar. Ağırlıkların uygulanmasıyla yansız tahminler elde edilecektir denilemez, ama yansızlığa çok yakın olunacağı söylenebilir (1, 10). Örnek araştırmalarında yardımcı değişken bilgisinin olması halinde, tahminlerin etkinliğini arttırmada ve varyansta da geniş çaplı azalmaların olmasını sağlamada bir ağırlıklandırma yöntemi olan ayarlama tahmin edicisinin kullanımı önerilebilir. Bu çalışmada ayarlanmış ağırlıkların nasıl hesaplandığı bir hipotetik veriyle gösterilmeye çalışılmıştır.

Araştırma verisinde yardımcı bilginin etkili kullanımı hem tahmin edilecek yığın değerlerine hem de ilgilenilen değişkeni ile yardımcı değişkenler arasındaki gerçek ilişkiye bağlıdır. Yardımcı değişkenler üzerinde bilinçsizce ayarlama yapmak çoğunlukla iyi bir yaklaşım değildir. Ayarlamada asıl düşünce ağırlıkları yeniden düzenleyerek, bilinen yardımcı değişken toplamlarını örnek için yeniden temsil edilir hale getirmektir.

Ayarlama tahmin edicilerinin yapılabilmesinde uzaklık ölçümü ve ayarlama denklemleri olarak adlandırılan iki temel bileşen vardır. Deville ve Särndal (1992)'ın önemli bir çıkarsaması da, yığın toplamı tahmin edicilerinin uzaklık ölçümünün farklı seçimlerinde tümünün asimptotik olarak denk olduklarını ve yığın toplamı için aynı asimptotik varyansa sahiptirler (11). Ayarlanmış ağırlıklarıyla yeni ağırlıklar, arzu edilen yansız tahminler sağlayan örnekleme tasarım ağırlıklarında ($d_i=1/\pi_i$ de) mümkün olduğunca az değişiklik yapılarak belirlenir. Araştırma istatistikçileri örnekleme tasarım ağırlıklarının yakınında durmayı isterler.

6. CONCLUSION

The surveyors believe that weights which give perfect estimates for the auxiliary variables should also give a good estimate for the interest variable. By application of the weights it cannot be said that unbiased estimators will be obtained, but the result will be rather close to the unbiasedness (1, 10). When auxiliary information is available in survey sampling, the use of calibration estimation which is called as a weighting method can be proposed to improve efficiency of estimates and to bring about a large variance reduction in variance. In this paper calculating the calibration weights were tried to be illustrated by a hypothetical data.

The effective use of auxiliary information in survey data depends on both the population value to be estimated and the actual correlation between interest variable and auxiliary variables. Usually assigning weights unconsciously on the auxiliary variables is not a good approach. The original idea in calibration is to modify the weights so that known auxiliary variable totals are reproduced for the sample.

In constructing the calibration estimators, there are two main components which are called as the distance measure and the calibration equation. An important result of Deville and Särndal (1992) is that, population total estimators with different choices of distance measure are all asymptotically equivalent and are all consistent for population total with the same asymptotic variance (11). The new weights are determined by the calibrated weights by making possibly little modifications on the sampling design weights ($d_i=1/\pi_i$) which provide desirable unbiased estimators. Survey statisticians wish to remain close to the sampling design weights.

KAYNAKLAR/ REFERENCES

1. Devılle, J. C. and Särndal, C. E., "Calibration estimators in survey sampling", ***Journal of the American Statistical Association***, 87: 376-382 (1992).

2. Kott, P. S., “On calibration weighting”, http://www.nass.usda.gov/research/reports/2003-jsm-kott.pdf (01.11.2003).

3. Betlehem, J. G. and Keller, W. J., “Linear weighting of the sample survey data”, ***Journal of Official Statistics***, Vol.3, 2: 141-153 (1987).

4. Kalton, G. and Flores-Cervantes, I., “Weighting methods”, ***Journal of the American Statistical Association***,Vol.19, 2: 81-97 (2003).

5. Wu, C. and Sitter, R.R., “A model-calibration approach to using complete auxiliary information from survey data”, ***Journal of the American Statistical Association***, 96: 5-193, (2001).

6. Montanari, G. E. and Ranalli M. G., “On calibration methods for design based finite population inferences”, https://www.isi2003.de/guest/IDbb1c3ff5c110b9/?MIval= AbstractView&ABSID=2460, (1.11.2003).

7. Creedy, J., Survey Reweighting for tax microsimulation modelling, New Zealand treasury working paper 03/17, http://www.treasury.govt.nz/workingpapers/2003/twp03-17.pdf (September 2003).

8. Estevao, V. M. and Särndal C. S, “A functional form approach to calibration”, ***Journal Of Official Statistics***, Vol.16, No.4, 379-399 (2000).

9. Särndal, C.E., Swensson, B., and Wretman, J., “Model assisted survey sampling”, ***Springer***, New York (1992).

10. Wu, C., “Optimal calibration estimators in survey sampling”, ***Biometrika***, **90**, to appear in the December 2003 issue, http://www.stats.uwaterloo.ca/~cbwu/paper.html, 20.10.2003.

11. Skinner, C., “Calibration weighting and non-sampling errors”, http://europa.eu.int/comm/ eurostat/research/index.htm?http://europa.eu.int/en/comm/eurostat/research/conferences/ntts-98/agenda.htm&1, 20.12.2004.

Received/ Geliş Tarihi: 31.03.2004   Accepted/Kabul Tarihi: 03.03.2005