HOW TO STANDARDIZE IF YOU MUST

MARCELLO D'AGOSTINO AND VALENTINO DARDANONI

How to Standardize if you Must

Marcello D'Agostino
Università di Ferrara

Valentino Dardanoni
Università di Palermo

Preliminary draft
Please do not circulate

```
   Standardize, tr.v.
To cause to conform to a standard.
To evaluate by comparing with a standard.
```
The American Heritage Dictionary of the English Language, Fourth Ed.

1 Introduction

In many situations one may wish to compare a given attribute or characteristic which refers to a certain context, with another attribute or characteristic referrring to a possibly different context. For example, one may wish to compare the grade of a student in a given school with the grade of another student in a different school, or the economic status of an individual who lives in a certain region, with that of another individual living in a different one. For example how does a $50.000 annual income in the US compare to a $10.000 annual income in Romania? Or, how does a 75/100 grade in mathematics in a good school compare to a 75/100 grade in a poor one? While such questions seem ubiquitous, we are not aware of any theoretical work on the "good ways" to accomplish this task. In this paper, we propose some properties that a standardization process should satisfy, and characterize a framework for being able to answer the questions posed above.

Suppose that you have a real valued vector $\mathbf{u}$ describing a given characteristic in a group of individuals in a given context $\mathbf{u}$, and another real valued vector $\mathbf{v}$ which measures the same characteristic for a group of individuals

in a different context $\mathbf{v}$. How can you compare u_i in $\mathbf{u}$ with v_j in $\mathbf{v}$? *Standardizing* means comparing u_i and v_j by conforming them to a standard context.[1] In this sense, our approach is akin to the comparison of demographic characteristics of a given population by transforming each population into a "standardized population" with given age and gender characteristics (xxx references).

Our approach is based on the search of good properties of a *standardization function*, which can be seen as a transformation of the elements of a vector into *standardized values*, i.e., elements belonging to a standard context. Our set of axioms allows us to characterize a class of standardization functions which includes some widely used procedures. As remarked above, while the transformation of real numbers into standardized values is a common endeavor in many applied and theoretical fields, we are not aware of previous literature on axiomatic characterizations of standardization.

Finally, we stress that our approach may be of use in the problem of multidimensional comparison of well-being, which is recently receiving a lot of theoretical, empirical and political attention. When comparing multidimensional attributes, standard practice suggests to aggregate different real valued well being indicators after somehow transforming each indicator into comparable units (see e.g. The Handbook on Constructing Composite Indicators, [**?**], for an extensive practical guide on the different data transformation typically employed).

2 Statistics

The first assumption we make is that data are represented by vectors of reals of size ≥ 2, so that our data space is a subset $\mathcal{V}$ of $\bigcup_{n=2}^{\infty} \mathbb{R}^n$. An element of $\mathcal{V}$ is a vector that you want to standardize.

Given a vector $\mathbf{u}$, how much information do we need about $\mathbf{u}_{-i}$ to be able standardize a given element u_i? On one extreme, we may insist that we must know *all* the elements of $\mathbf{u}_{-i}$. On the other, we may think that we need

[1] Probably the most famous standardization procedure -namely the creation of standardized values in statistics- can be seen as an application of this principle; since computing normal probabilities with arbitrary mean and variance traditionally required a certain degree of effort, it made sense to compare values under different normal distributions by comparing them with a standard (zero mean unit variance) one for which tables were calculated.

no information about $\mathbf{u}_{-i}$, in which case the very notion of standardization is meaningless. In most situations, however, we may be happy with partial information in terms of *statistics*.

Definition 1 *A* statistic *over $\mathcal{V}$ is a permutation invariant function $\tau : \mathcal{V} \mapsto \mathbb{R}$. Given a set of statistics $\mathcal{T}$ over $\mathcal{V}$, we say that a statistic τ is* primitive in $\mathcal{T}$ *if (i) $\tau \in \mathcal{T}$ and (ii) there are no finite sequence $\tau_1, \ldots, \tau_k \in \mathcal{T}$ and no function f such that, for all $\mathbf{x} \in \mathcal{V}$, $\tau(\mathbf{x}) = f(\tau_1(\mathbf{x}), \ldots, \tau_k(\mathbf{x}))$.*

Definition 2 *A statistic τ is*

- homogeneous *(of the first degree) if $\tau(a\mathbf{u}) = a\tau(\mathbf{u})$ for every vector $\mathbf{u} \in \mathcal{V}$ and every scalar $a > 0$;*
- weakly additive *if $\tau(\mathbf{u} + b) = \tau(\mathbf{u}) + \tau(\mathbf{b})$ for every vector $\mathbf{u} \in \mathcal{V}$ and every scalar b, where $\mathbf{b}$ is the vector in $\mathbb{R}^{|\mathbf{u}|}$ such that all its elements are equal to b.*
- *a* location statistic *if $\tau(\mathbf{u} + a) = \tau(\mathbf{u}) + a$ for every vector $\mathbf{u}$ and every scalar a;*
- *a* dispersion statistic *if $\tau(\mathbf{u} + a) = \tau(\mathbf{u})$ for every vector $\mathbf{u}$ and every scalar a.*

We shall assume that there is always a fixed *finite amount of statistical information* about elements of $\mathcal{V}$ (no matter how large they are), which is *sufficient* to make all the necessary discriminations required for standardization purposes. Such fixed set of summarizing statistics, call it $\mathcal{T} = \{\tau_1, \ldots, \tau_k\}$, is used as a "sieve" in order to filter out the irrelevant information. In other words, $\mathcal{T}$ is a sufficient set of statistics for standardization purposes.

3 Standardization frame

Let $\sim_{\mathcal{T}}$ be the equivalence relation defined as follows: $\mathbf{x} \sim_{\mathcal{T}} \mathbf{y}$ if and only if $\tau(\mathbf{x}) = \tau(\mathbf{y})$ for all $\tau \in \mathcal{T}$. In other words, each equivalence class induced by $\sim_{\mathcal{T}}$ consists in the set of all solutions of the following system of equations:

$$\begin{aligned} \tau_1(\mathbf{x}) &= c_1 \\ &\vdots \\ \tau_n(\mathbf{x}) &= c_n, \end{aligned}$$

where $\tau_1, \ldots, \tau_n$ are all the elements of $\mathcal{T}$ and $c_1, \ldots, c_n$ are constants.

Standardization then depends on the partition of $\mathcal{V}$ induced by the statistical information conveyed by the chosen set $\mathcal{T}$: the larger the amount of information, the finer the partition, and $\mathcal{T}$-equivalent vectors will be treated exactly in the same way for standardization purposes. In other words, we maintain that the "context" relative to which the standardized value is determined is not a vector, but an *equivalence class* of vectors identified by the values of the chosen statistics.

Definition 3 *A* standardization frame *is a triple* $\mathcal{F} = \langle \mathcal{V}, \mathcal{T}, D \rangle$ *such that:*

1. $\mathcal{T}$ *is a finite set of statistics over* $\mathcal{V}$ *such that for every* $\tau \in \mathcal{T}$, τ *is primitive in* $\mathcal{T}$*;*

2. D *is a distinguished element of* $\mathcal{V}/\sim_{\mathcal{T}}$*, i.e. a distinguished equivalence class in the partition of* $\mathcal{V}$ *induced by* $\sim_{\mathcal{T}}$*.*

In the sequel, we will assume that all statistics in $\mathcal{T}$ are homogeneous of the first degree and weakly additive. This is not a severe restriction since most statistics one may want to use in this context either enjoy the two properties above (such as mean, mode, quantiles, maximum and minimum, standard deviation, interquantile dispersion) or may be expressed as functions of statistics which enjoy them (such as variance, skewness, kurtosis, moments, positive power means).[2]

Definition 4 *We say that* $\mathcal{T}$ *is XXX if and only for every* $E \in \mathcal{V}/\sim_{\mathcal{T}}$, $\mathrm{dom}(E)$ *is an interval, where* $\mathrm{dom}(E)$ *is defined as follows:*

$$\mathrm{dom}(E) = \{x | \exists \mathbf{z} \in E, x \textit{ occurs in } \mathbf{z}\};$$

4 Standardization function

We come now to main purpose of this paper, that is, to investigate a *standardization function* over the standardization frame $\mathcal{F}$.

[2] To mention just one example, the m-order moment $\frac{1}{n}\sum_{i=1}^{n} x_i^m$ can be expressed as a function of the mean $\mu(\mathbf{x})$ and the set of statistics $(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu(\mathbf{x}))^j)^{1/j}$ for $j = 2, \cdots, m$, which enjoy our two properties.

Definition 5 *Given a standardization frame $\mathcal{F}$, a* standardization function *over $\mathcal{F}$ is a mapping S from $\mathbb{R} \times \mathcal{V}$ into $\mathbb{R}$, which is strictly increasing in the first argument.*

Let $S(\mathbf{x})$ denote the vector obtained from $\mathbf{x}$ by applying the standardization function S to each of its elements, that is, $S(\mathbf{x}) = [S(x_1, \mathbf{x}), \ldots, S(x_n, \mathbf{x})]$.

Consider the following properties that S may satisfy:

Property 1 *$S(\mathbf{x}) \in D$ for every $\mathbf{x} \in \mathcal{V}$.*

Property 1 defines a standardization function as a function which transforms data into a standard context (i.e., S maps any vector in $\mathcal{V}$ into one in the distinguished class D), which is the key property of the standardization function.

From this property and the definition of a standardization frame it follows that, for every $\mathcal{F} = \langle \mathcal{V}, \mathcal{T}, D \rangle$ and every S over $\mathcal{F}$,

$$\tau(S(\mathbf{u})) = \tau(S(\mathbf{v})) \text{ for every } \tau \in \mathcal{T} \text{ and } \mathbf{u}, \mathbf{v} \in \mathcal{V}. \tag{1}$$

Property 2 *$S(x_i, \mathbf{x}) = x_i$ whenever $\mathbf{x} \in D$.*

Property 2 says that when a given vector is already in the distinguished class D, we are happy to leave it unchanged. Notice that, in conjunction with Property 1, this property implies that S is *stable*, that is, $S(S(\mathbf{x})) = S(\mathbf{x})$ for every $\mathbf{x} \in \mathcal{V}$.

Property 3 *Whenever $x_i = v_j$ and $\mathbf{x} \sim_{\mathcal{T}} \mathbf{v}$, $S(x_i, \mathbf{x}) = S(v_j, \mathbf{v})$.*

Property 3 basically expresses the fact that $S(x_i, \mathbf{x})$ is a function of x_i and of the equivalence class of $\mathbf{x}$.

Property 4 *for every $E \in \mathcal{V}/\equiv_{\mathcal{T}}$, every $\mathbf{u}, \mathbf{v}, \mathbf{w} \in E$ and every $x \in \mathbf{u}$, $y \in \mathbf{v}$, $z \in \mathbf{w}$,*

$$|x - y| \leq |x - z| \text{ if and only if } |S(x, \mathbf{u}) - S(y, \mathbf{v})| \leq |S(x, \mathbf{u}) - S(z, \mathbf{w})|.$$

This last property says that S preserves the relation "x is closer to y than to z" between elements of equivalent contexts $\mathbf{u}, \mathbf{v}, \mathbf{z}$. For example, it says that if Joe, Matt and Jane belong to equivalent classes (with respect to the given statistics) and Joe's grade in math is closer than Matt's to Jane's before the standardization, then it should also be closer after standardization.

5 Result

Theorem 1 *Let S be a standardization function over $\mathcal{F} = \langle \mathcal{V}, \mathcal{T}, D \rangle$ such that $\#\mathcal{T} > 1$ and $\mathcal{T}$ is XXX. S satisfies Properties 1–4 if, and only if, for every non-degenerate*[3] *$\mathbf{u} \in \mathcal{V}$,*

$$S(u_i, \mathbf{u}) = c \cdot \big(\frac{u_i - \tau_2(\mathbf{u})}{\tau_1(\mathbf{u})}\big) + d$$

for some dispersion statistic τ_1 and location statistic τ_2, where the constants c and d equal respectively to the value of τ_1 and τ_2 in the distinguished class D.

The theorem basically says that whenever S satisfies Properties 1–4:

1. No matter how many statistics we may think it appropriate to use for standardization purposes, the set $\mathcal{T}$ must contain exactly two statistics;
2. These two statistics are fairly precisely defined: one must be a location statistic and one a dispersion statistic;
3. Once choice is made of the favourite location and dispersion statistics and their value in the distinguished equivalence class, the functional from of S is exacly determined.

Of course the theorem characterizes two very common standardization procedures: if the mean and the standard deviation are chosen as favourite location and dispersion statistics, and D is the set of vectors with zero mean and unitary standard deviation,

$$S(u_i, \mathbf{u}) = \frac{u_i - \mu(\mathbf{u})}{\sigma(\mathbf{u})},$$

while if the minimum and the range are chosen as favourite location and dispersion statistics, and D is the set of vectors $\mathbf{x} \in D$ such that $min(\mathbf{x}) = 0$ and $max(\mathbf{x}) - min(\mathbf{x}) = 1$,

$$S(u_i, \mathbf{u}) = \frac{u_i - min(\mathbf{u})}{max(\mathbf{u}) - min(\mathbf{u})}.$$

On the other hand, the theorem also tells that other common standardization procedures must violate some of our properties. For example, using ranks for standardization violates Property 4.

[3] We say that a vector is *degenerate* when all of its elements are equal.

6 Appendix: Proof of the Theorem

To prove the theorem, we first need two preliminary lemmas.

Lemma 1 *Let $\mathcal{F} = \langle \mathcal{V}, \mathcal{T}, D \rangle$ be a standardization frame. We can assume without loss of generality that every statistic $\tau \in \mathcal{T}$ is either a dispersion or a location statistic.*

Proof First, observe that a weakly additive statistic is: (i) a location statistic if and only if $\tau(\mathbf{a}) = a$ for every scalar a and every vector $\mathbf{a}$ such that all its elements are equal to a; (ii) a dispersion statistic if and only if $\tau(\mathbf{a}) = 0$ for every scalar a and every vector $\mathbf{a}$ such that all its elements are equal to a.

Let now τ be any statistic in $\mathcal{T}$. Let $\mathbf{1}$ be any vector such that all its elements are equal to 1 and $\mathbf{a}$ be any vector of the same size as $\mathbf{1}$ and such that all its elements are equal to a. There are two cases: if $\tau(\mathbf{1}) = 0$, then τ is a dispersion statistic, since we assume that all statistics are homogeneous of the first degree and, therefore, $\tau(\mathbf{a}) = a\tau(\mathbf{1}) = 0$ (recall that all statistics in $\mathcal{T}$ are assumed to be weakly additive and the latter identity is a necessary and sufficient condition for any weakly additive statistic to be a dispersion statistic); on the other hand, if $\tau(\mathbf{1}) = c \neq 0$, then $\tau(\mathbf{a}) = a\tau(\mathbf{1}) = ca$ and so $\frac{1}{c}\tau(\mathbf{a}) = a$. Therefore, $\frac{1}{c}\tau$ is a location statistic, since the latter identity is a necessary and sufficient condition for a weakly additive statistic to be a location statistic. Now, since $\frac{1}{c}\tau(\mathbf{x})$ is a function of $\tau(\mathbf{x})$, the standardization frame $\mathcal{F}'\langle \mathcal{V}, \mathcal{T}', D \rangle$, where $\mathcal{T}'$ is the set of statistics obtained from $\mathcal{T}$ by replacing τ with $\frac{1}{c}\tau$ induces the same partition on $\mathcal{V}$ as the original frame. So every standardization function over $\mathcal{F}$ is also a standardization function over $\mathcal{F}'$. □

Lemma 2 *For every standardization frame $\mathcal{F} = \langle \mathcal{V}, \mathcal{T}, D \rangle$, every standardization function S over $\mathcal{F}$, every equivalence class $E \in \mathcal{V}/\equiv_{\mathcal{T}}$ and every $\mathbf{u} \in E$:*

$$S(u_i, \mathbf{u}) = a_E u_i + b_E,$$

for some $a_E > 0$ and b_E depending only on E.

Proof

Let E be an arbitrary equivalence class in $\mathcal{V}/\equiv_{\mathcal{T}}$ and x and y be two arbitrary real numbers in $\mathrm{dom}(E)$ such that $x + y, x - y \in \mathrm{dom}(E)$. Now, observe that it follows from Property 4 and Definition 5 that

$$S(x + y, \mathbf{u}) - S(x, \mathbf{v}) = S(x, \mathbf{v}) - S(x - y, \mathbf{w}),$$

for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in E$ such that $x+y$ occurs in $\mathbf{u}$, x occurs in $\mathbf{v}$ and $x-y$ occurs in $\mathbf{w}$. Notice that, within the same equivalence class E, the function S is indepedent of its second argument (by Property 3), since the value of S is preserved under substitutions of the vector in the second argument with an equivalent one. So, let $S_E(z)$ be the function defined as follows: $S_E(z) = x$ if and only if $S(z, \mathbf{w}) = x$ for every vector $\mathbf{w} \in E$ such that $\mathbf{w}$ contains z. This function is defined for every real in $\mathrm{dom}(E)$. The argument above shows that:

$$S_E(x+y) - S_E(x) = S_E(x) - S_E(x-y),$$

for all x and y in $\mathrm{dom}(E)$. Therefore S_E must be a linear function, that is, there exist constants a_E and b_E (with $a_E > 0$ since S is strictly increasing) such that:

$$S_E(z) = a_E z + b_E.$$

This concludes the proof of the lemma. □

We are now ready to prove the theorem. It is easy to see that any S taking the assumed form satisfies Properties 1–4. To prove the converse, let us assume that S is a standardization function over $\mathcal{F} = \langle \mathcal{V}, \mathcal{T}, D \rangle$ and assume that the statistics in $\mathcal{T}$ have been ordered in some arbitrary way, that is $\mathcal{T} = \{\tau_1, \ldots, \tau_k\}$, $k > 1$.

Recalling that, by Lemma 2, $S(u_i, \mathbf{u}) = a_E u_i + b_E$ for some a_E, b_E depending only on E let α and β be functions $\mathbb{R}^k \mapsto \mathbb{R}$ such that $\alpha(\tau_1(\mathbf{u}), \ldots, \tau_k(\mathbf{u})) = a_E$ and $\beta(\tau_1(\mathbf{u}), \ldots, \tau_k(\mathbf{u})) = b$ for every $E \in \mathcal{V}/\equiv$ and every $\mathbf{u} \in E$. Hence, we have that:

$$S(u_i, \mathbf{u}) = \alpha_{\mathcal{T}}(\mathbf{u}) u_i + \beta_{\mathcal{T}}(\mathbf{u})),$$

where

$$\alpha_{\mathcal{T}}(\mathbf{u}) = \alpha(\tau_1(\mathbf{u}), \ldots, \tau_k(\mathbf{u}))$$

and

$$\beta_{\mathcal{T}}(\mathbf{u}) = \beta(\tau_1(\mathbf{u}), \ldots, \tau_k(\mathbf{u})).$$

Given Lemma 1, we may distinguish two cases.

Case 1. The first statistic in $\mathcal{T}$ is a dispersion statistic, say δ. Since δ is homogeneous of the first degree and weakly additive, then, by equation 1,

for some constant c:

$$\begin{aligned} \delta\big(S(\mathbf{u})\big) = \delta(\alpha_{\mathcal{T}}(\mathbf{u})\mathbf{u} + \beta_{\mathcal{T}}(\mathbf{u})) &= \delta(\alpha_{\mathcal{T}}(\mathbf{u})\mathbf{u}) + \delta(\beta_{\mathcal{T}}(\mathbf{u})) \\ &= \alpha_{\mathcal{T}}(\mathbf{u})\delta(\mathbf{u}) \\ &= c. \end{aligned} \tag{2}$$

Hence $\alpha_{\mathcal{T}}(\mathbf{u}) = c/\delta(\mathbf{u})$.

Observe that, if δ' is another dispersion statistic in $\mathcal{T}$, then $\alpha_{\mathcal{T}}(\mathbf{u}) = c'/\delta'(\mathbf{u})$ for some constant c' and so $\delta'(\mathbf{u}) = \frac{c'}{c}\delta(\mathbf{u})$. But δ' is a function of δ. So, since $\mathcal{T}$ contains only statistics which are primitive in it, and $\delta \in \mathcal{T}$, then $\delta' \notin \mathcal{T}$. Therefore, $\mathcal{T}$ cannot contain any other dispersion statistic.

Therefore, by Lemma 1, the next statistic in $\mathcal{T}$, must be a location statistic. Let us call it λ. Again, since λ is homogeneous of the first degree and weakly additive, we have that, by equation 1:

$$\begin{aligned} \lambda\big(S(\mathbf{u})\big) = \lambda(\alpha_{\mathcal{T}}(\mathbf{u})\mathbf{u} + \beta_{\mathcal{T}}(\mathbf{u})) &= \lambda(\alpha_{\mathcal{T}}(\mathbf{u})\mathbf{u}) + \lambda(\beta_{\mathcal{T}}(\mathbf{u})) \\ &= \alpha_{\mathcal{T}}(\mathbf{u})\lambda(\mathbf{u}) + \beta_{\mathcal{T}}(\mathbf{u}) \\ &= d \end{aligned} \tag{3}$$

for some constant d.

Hence, $\beta_{\mathcal{T}}(\mathbf{u}) = d - \alpha_{\mathcal{T}}(\mathbf{u})\lambda(\mathbf{u})$ and, replacing $\alpha_{\mathcal{T}}(\mathbf{u})$ with its value obtained above, $\beta_{\mathcal{T}}(\mathbf{u}) = d - \frac{c}{\delta(\mathbf{u})}\lambda(\mathbf{u})$. So:

$$S(u_i, \mathbf{u}) = \alpha_{\mathcal{T}}(\mathbf{u})u_i + \beta_{\mathcal{T}}(\mathbf{u}) = \frac{c(u_i - \lambda(\mathbf{u}))}{\delta(\mathbf{u})} + d.$$

Since at this stage S is completely determined, any other statistic in $\mathcal{T}$ is redundant.

Case 2: The first statistic in $\mathcal{T}$ is a location statistic. Let us call it λ. Then, since λ is homogeneous of the first degree and weakly additive, we have, as in (3) above, that $\beta_{\mathcal{T}}(\mathbf{u}) = d - \alpha_{\mathcal{T}}(\mathbf{u})\lambda(\mathbf{u})$. Let us, then, try to determine $\alpha_{\mathcal{T}}(\mathbf{u})$. Now, if the next statistic in $\mathcal{T}$ is a dispersion statistic, then by (2) above, we fall back to case 1. Suppose, then, that the next statistic in $\mathcal{T}$ is a location statistic, call it λ'. Then, using again the equations (3) above, with λ' instead of λ and a new constant e instead of d, we obtain $\beta_{\mathcal{T}}(\mathbf{u}) = e - \alpha_{\mathcal{T}}(\mathbf{u})\lambda'(\mathbf{u})$. So

$$d - \alpha_{\mathcal{T}}(\mathbf{u})\lambda(\mathbf{u}) = e - \alpha_{\mathcal{T}}(\mathbf{u})\lambda'(\mathbf{u})$$

and

$$\alpha_{\mathcal{T}}(\mathbf{u}) = \frac{e-d}{\lambda'(\mathbf{u}) - \lambda(\mathbf{u})}.$$

Now, observe that, since λ' and λ are location statistics, their difference is a dispersion statistic, for

$$\begin{aligned} \lambda(\mathbf{u}+\mathbf{a}) - \lambda'(\mathbf{u}+\mathbf{a}) &= \lambda(\mathbf{u}) + \lambda(\mathbf{a}) - \lambda'(\mathbf{u}) - \lambda'(\mathbf{a}) \\ &= \lambda(\mathbf{u}) + a - \lambda'(\mathbf{u}) - a \\ &= \lambda(\mathbf{u}) - \lambda'(\mathbf{u}) \end{aligned} \tag{4}$$

Let $\delta(\mathbf{u}) = \lambda'(\mathbf{u}) - \lambda(\mathbf{u})$ and let $c = e - d$, then we have again that:

$$S(u_i, \mathbf{u}) = \alpha_{\mathcal{T}}(\mathbf{u})u_i + \beta_{\mathcal{T}}(\mathbf{u}) = \frac{c(u_i - \lambda(\mathbf{u}))}{\delta(\mathbf{u})} + d.$$

Since S is again completely determined, any other statistic in $\mathcal{T}$ is redundant.

Notice now that, using the fact that τ_1 and τ_2 are respectively a dispersion and a location statistic, $\tau_1(S(\mathbf{x})) = c\frac{\tau_1(\mathbf{x})}{\tau_1(\mathbf{x})} = c$ and $\tau_2(S(\mathbf{x})) = c(\frac{\tau_2(\mathbf{x})}{\tau_1(\mathbf{x})} - \frac{\tau_2(\mathbf{x})}{\tau_1(\mathbf{x})}) + d = d$ for any $\mathbf{x} \in \mathcal{V}$. Using Property 1 and equation 1 concludes the proof of the theorem.