A Theory of Dichotomous Valuation with Applications to Variable Selection

Hu, Xingwei

International Monetary Fund

15 June 2017

A Theory of Dichotomous Valuation with Applications to Variable Selection

Xingwei Hu[1]

International Monetary Fund, 700 19th St NW, Washington, DC 20431, USA

Abstract

An econometric or statistical model may undergo a marginal gain when a new variable is admitted, and marginal loss if an existing variable is removed. The value of a variable to the model is quantified by its expected marginal gain and marginal loss. Under a prior belief that all candidate variables should be treated fairly, we derive a few formulas which evaluate the overall performance of each variable. One formula is identical to that for the Shapley value. However, it is not symmetric with respect to marginal gain and marginal loss; moreover, the Shapley value favors the latter. Thus we propose a unbiased solution. Two empirical studies are included: the first being a multi-criteria model selection for a dynamic panel regression; the second being an analysis of effect on hourly wage given by additional years of schooling.

Keywords: `unbiased multivariate Shapley value`, `variable selection`, `marginal effect`, `endowment bias`, `model uncertainty`

JEL Classification Number: C11, C52, C57, C71, D81

1. Introduction

The problem with which we are concerned relates to the following typical situation: When modeling data, we attempt to use a formula to simplify the

Email address: `xhu@imf.org` (Xingwei Hu)

[1]The author thanks Lloyd S. Shapley, David F. Hendry, Stephen Legrand, Michael A. Jones, Shiquan Wu, and Alisha Abdelilah for their suggestions and thanks Alan Krueger and Cecilia Rouse for the Twinsburg schooling dataset used in the empirical study.

underlying process that has been operating to produce the data. To be useful, the model should not only help us to better understand the underlying structure of the variables in the past, but also be predictive in a non-specific situation in the future; ideally, it should perform well under multiple criteria. In general, the variables may either correlate with or be interdependent on each other. Some variables may even be explained by other variables; thus, they are highly superfluous. Others may simply be irrelevant to the underlying process. Selection of right variables in modeling and forecasting the underlying process is one of the most fundamental problems in statistics and econometrics.

In this paper, we incorporate rather than ignore model uncertainty and derive a few formulas to be used in variable selection, by evaluating the contribution of each variable in a large set of modeling scenarios . The exposition is self-contained and the approach employed is game-theoretic and Bayesian. For a simple context, let Y denote the dependent variables in the model, and let the universal of candidate explanatory variables be indexed by $N = \{1, 2, \cdots, n\}$. We use "$\bar{i}$" for the singleton set $\{i\}$ and "$\backslash$" for set subtraction. For any set $T \subseteq N$, let $|T|$ denote the cardinality of T, and let $v(T)$ be a vector of collective performance measures when we model the data Y using all variables in T. In particular, for the empty set $\emptyset$, $v(\emptyset)$ is the performance measures when the model does not involve any variables in N; say, for example, when modeling Y by a constant or a linear trend. The performance measures could be a measure of model fit, variance explained, predictive power, negative of the cost function, probability of avoiding fatal errors, or any combination of the above.

For a given performance function $v : 2^N \to R^m$, our goal is to find a small set of variables that have high importance with respect to v. Unlike many algorithm-based approaches, which search for a subset of variables with optimal collective performance, we directly evaluate the *overall performance* of each variable under a fair prior belief. Then, we use their performance to select variables and institute the model. The search for a single model, however, ignores the model uncertainty. By the term "overall performance," we mean the average performance in all possible modeling scenarios, not in a specific model. Thus, a prior belief or distribution is required to specify the possibility of modeling scenarios; in this sense, our approach is also Bayesian and uses model averaging (e.g., George and McCulloch, 1997; Clyde and George, 2004; O'Hara and Sillanpaa, 2009). Given special classes of priors, we show that the expected performance coincides with the Shapley value and the Banzhaf value in the coalitional game (N, v). We also suggest a new value concept which embraces both the Shapley value and the Banzhaf value.

The results in this paper arise from considering any marginal effect to be either a marginal loss or a marginal gain. As an example, consider the contribution of a bachelor's degree (hereinafter "BD") to the annual income of an individual aged 40. The individual may have a BD or not. For an individual with a BD, a marginal gain is computed as the difference between his current annual income and his estimated annual income, assuming he had no BD, *ceteris paribus*. On the other hand, a marginal loss for an individual with no BD is computed as the difference between his current annual income and his estimated annual income, assuming he had a BD, *ceteris paribus*. We note that the possession of a BD is interwoven with other factors, such as his profession and length of relevant work experience, which also affect his income. We define the value of a BD as the expected marginal gain and marginal loss, incorporating the ownership uncertainty and the interdependence with other factors. In addition, a BD holder may show some endowment bias, valuing the BD more than those who don't have a BD; we define the bias as the expected difference between the marginal gain and loss.

Our research here is closely related with the Shapley value (Shapley, 1953). Recently, the use of the Shapley value in modeling data has gained popularity (e.g., Lipovetsky and Conklin, 2001; Israeli, 2007; Gromping, 2007; Devicienti, 2010; Budina et al., 2015), partly owing to its simplicity and generality. The vast literature also provides us with many variations on the Shapley value (cf Donderer and Samet, 2002; Winter, 2002). Our research here provides not only a new proof and a new interpretation of the Shapley value, but also a theoretical foundation for properly using the value concept in variable selection and related fields. Theorem 1 in this paper was previously proved in Hu(2002); the first two theorems of this paper are a generalization of the results in Hu (2006) in which the grand cooperation is not necessary and $v(T)$ is a binary function representing a vote passing or blocking by the voters of T.

The advantage of our approach is fivefold. First, the performance measurement v is a vector function and thus extends the Shapley value to the multivariate case. As a consequence, the variable selection eventually becomes a multi-criteria decision analysis, mitigating the discrepancy between model accuracy and usefulness. Secondly, we acknowledge the model uncertainty and the inter-linking among the variables; we address them using a prior distribution or prior belief. Any further specifications or restrictions can also be added to the priors. Thirdly, by dichotomizing the marginal effect, we generalize the Shapley value and the Banzhaf value under a same framework; they differ in prior uniform distributions. Next, we discover the symmetry in the Banzhaf value and asymmetry in the Shapley value and suggest a simple adjustment to be used in the applications. Lastly, under the same framework we introduce a new valuation solution based on binomial distributions. It is as tractable in expression and calculation as the Shapley value and the Banzhaf value; additionally, it allows the expected model size from the prior to be consistent with the model being estimated. Moreover, it has a constant endowment bias ratio and relates to all other value concepts discussed in this paper.

The paper is organized as follows. Section 2 formulates the essential ideas of dichotomized marginal effect and relate them to the Shapley value and the Banzhaf value. Next, sections 3 analyzes the difference, called *endowment bias*, between the expectations of marginal gain and marginal loss. In this section, a weighted unbiased value concept is proposed. After that, section 4 discusses how to implement the basic ideas by a sequential algorithm. In section 5, we conduct two empirical studies. In section 6, we relate this framework to several other topics in economics, finance, and statistics. Proofs and large tables are contained in the Appendix.

2. Evaluation of Candidate Predictors

To address the model uncertainty, let the random set $\mathbf{S} \subseteq N$ be the set of variables in the model. The randomness of $\mathbf{S}$ arises from both objective uncertainty and subjective one. For the objective uncertainty, an econometric or statistical model is merely an approximation and simplification of reality − Box (1976) even claimed that all models are wrong − and there could be many good approximations under various criteria. For the subjective uncertainty, we do not know exactly which specific variables to choose before performing a selection analysis; we may have a class of subjective probabilities for it, derived from our personal judgment, opinions, past experience, or even fairness assumptions.

Let μ be the distribution of $\mathbf{S}$ with P_T being the probability of $\mathbf{S} = T$. Without any specific prior knowledge, we have no reason to believe that one set of candidate variables is more likely to be $\mathbf{S}$ than another set of the same size. That is, we should not discriminate between sets of variables having the same size. To put it in another way, given the size of $\mathbf{S}$, we have no reason to select one variable and reject another. This argument is formally justified by the *principle of insufficient reason* and defines a class of distributions:

$$\mathcal{F} \stackrel{\text{def}}{=} \{\mu | P_T \text{ is a function depending only on the size of } T\}.$$

For the indeterminate $\mathbf{S}$, we could add one variable from $N \setminus \mathbf{S}$ to the model or we could remove an existing variable from the model. The worth of the addition or removal can be explained by the variable's *marginal effect* to $v(\mathbf{S})$. There are two different scenarios:

- Scenario 1: $i \in \mathbf{S}$. Then, i's marginal effect is $v(\mathbf{S}) - v(\mathbf{S} \setminus \bar{i})$, called *marginal gain*, in that it contributes $v(\mathbf{S}) - v(\mathbf{S} \setminus \bar{i})$ when we include it into the model. The expected marginal gain, due to i's participation in $\mathbf{S}$, is

$$\gamma_i[v; \mu] \stackrel{\text{def}}{=} \mathrm{E}\left[v(\mathbf{S}) - v(\mathbf{S} \setminus \bar{i})\right].$$

- Scenario 2: $i \notin \mathbf{S}$. Then, the marginal effect is $v(\mathbf{S} \cup \bar{i}) - v(\mathbf{S})$ in that $\mathbf{S}$ faces a *marginal loss* or opportunity cost $v(\mathbf{S} \cup \bar{i}) - v(\mathbf{S})$ without variable i in the model. In other words, the ith variable could have increased the collective performance by $v(\mathbf{S} \cup \bar{i}) - v(\mathbf{S})$ if we had added it to $\mathbf{S}$. The expected marginal loss, due to i's absence from $\mathbf{S}$, is then

$$\lambda_i[v; \mu] \stackrel{\text{def}}{=} \mathrm{E}\left[v(\mathbf{S} \cup \bar{i}) - v(\mathbf{S})\right].$$

In either case, the marginal effect of i to $v(\mathbf{S})$ can be written as $v(\mathbf{S} \cup \bar{i}) - v(\mathbf{S} \setminus \bar{i})$. Combining these marginals, we define variable i's *dichotomous value* $\psi_i[v; \mu]$ (hereinafter, "*D-value*"), as a functional function of v, by its expected marginal effect under the distribution μ:

$$\psi_i[v; \mu] \stackrel{\text{def}}{=} \gamma_i[v; \mu] + \lambda_i[v; \mu]. \tag{1}$$

Equivalently,

$$\psi_i[v; \mu] \stackrel{\text{def}}{=} \mathrm{E}\left[v(\mathbf{S} \cup \bar{i}) - v(\mathbf{S} \setminus \bar{i})\right]. \tag{2}$$

In other words, the D-value $\psi_i[v; \mu]$ quantifies variable i's overall performance in v under the distribution μ, which itself specifies the likelihood of modeling scenarios. Formally, one could derive (2) from two axioms:

- Marginality: given $P_{T} = 1$, $\psi_i[v; \mu] = v(T \cup \bar{i}) - v(T \setminus \bar{i})$;
- Linearity: for any μ_1, μ_2 on 2^N and any $0 \leq \alpha \leq 1$,

$$\psi[v; \alpha\mu_1 + (1 - \alpha)\mu_2] = \alpha\psi[v; \mu_1] + (1 - \alpha)\psi[v; \mu_2].$$

Besides, one of our objectives is to apply the D-value $\psi[v; \mu]$ to distribute the total collective performance $v(N) - v(\emptyset)$ among all candidate variables; this requires that μ satisfies the functional equation of

$$\sum_{i=1}^{n} \psi_i[v;\mu] = v(N) - v(\emptyset). \tag{3}$$

Thus, the portion $\psi_i[v;\mu]$ of $v(N) - v(\emptyset)$ is explained by variable i. The following theorem relates the D-value $\psi[v;\mu]$ with the Shapley value $\Psi[v]$ in the coalitional game (N, v), defined as:

$$\Psi_i[v] \stackrel{\text{def}}{=} \sum_{T \subseteq N} \frac{(|T|-1)!(n-|T|)!}{n!} \left[v(T) - v(T \setminus \bar{i}) \right].$$

Theorem 1. *For any $\mu \in \mathcal{F}$ which satisfies (3), we have $\psi[v;\mu] = \Psi[v]$ and*

$$P_T = \frac{(|T|)!(n-|T|)!}{(n+1)!} + (-1)^{|T|}\eta \tag{4}$$

for any η which guarantees the non-negativity of all P_T's.

Corollary 1. *If P_T is as in (4), then $\mu \in \mathcal{F}$ and $\psi[v;\mu] = \Psi[v]$.*

In deriving the proofs, we assume neither $v(\emptyset) = 0$, nor monotonicity of v, nor superadditivity of v (i.e., $v(X \cup Y) \geq v(X) + v(Y)$ for any disjoint X and Y in N). As shown by (4), μ is not even unique. Besides, in the benchmark prior with $\eta = 0$, $|\mathbf{S}|$ has the uniform distribution on the integers $0, 1, \cdots, n$.

In the game theory literature, there is another value concept called the Banzhzf value $b[v]$ (Banzhaf, 1965) which is defined as

$$b_i[v] \stackrel{\text{def}}{=} \frac{1}{2^{n-1}} \sum_{T} \left[v(T) - v(T \setminus \bar{i}) \right].$$

In the next theorem, we associate the Banzhaf value with the D-value $\psi[v;\mu]$ through another benchmark distribution, the uniform distribution on 2^N.

Theorem 2. *$\psi[v;\mu] = b[v]$ if and only if $P_T = \frac{1}{2^n} + (-1)^{|T|}\eta$ for some η with $|\eta| \leq \frac{1}{2^n}$.*

In this benchmark distribution (with $\eta = 0$), the model size $|\mathbf{S}|$ has a binomial distribution with parameters n and .5. Note that in both benchmark distributions, the model $\mathbf{S}$ has an expected size $\frac{n}{2}$, which could be highly unrealistic, especially when applied in big data analytics. To solve the puzzle, we may use priors with nonzero η in which the density function P_{T} oscillates with the size $|T|$; we could also consider other distributions for $|\mathbf{S}|$.

Theorem 3. *If $\mu \in \mathcal{F}$ and $|\mathbf{S}|$ has a distribution of Binomial(n,p), then*

$$P_{T} = p^{|T|}(1-p)^{n-|T|}$$

and

$$\begin{aligned} \psi_i[v;\mu] &= \tfrac{1}{p}\gamma_i[v;\mu] = \tfrac{1}{1-p}\lambda_i[v;\mu] \\ &= \sum_T p^{|T|-1}(1-p)^{n-|T|}\left[v(T) - v(T \setminus \bar{i})\right]. \end{aligned} \tag{5}$$

In the binomial distribution, the model $\mathbf{S}$ has an expected size of np.

3. Endowment Bias

In our valuation paradigm, we classify two types of marginal effect according to the ownership: $i \in \mathbf{S}$ or $i \notin \mathbf{S}$. In practice, people tend to value more on things they own, rather than ones they do not own $-$ even when things are exchangeable in value. This subjective bias is called the *endowment effect* or *endowment bias*. In this section, we first define the bias and then analyze the implied bias in the Shapley value and the Banzhf value.

For i, we define its *endowment bias* as the difference between its expected marginal gain and its expected marginal loss,

$$\kappa_i[v;\mu] \stackrel{\text{def}}{=} \gamma_i[v;\mu] - \lambda_i[v;\mu].$$

Lemma 1 provides a method to calculate biases directly without involving the expected marginal gain or loss.

Lemma 1. $\kappa_i[v;\mu] = \sum_T \left[2P_{T} - P_{T\cup\bar{i}} - P_{T\setminus\bar{i}}\right] v(T)$.

In Lemma 1, the weights of $v(T)$'s sum to zero; but the bias itself may be not. The following two theorems indicate strong association between the endowment unbiasedness and the Banzhaf value.

Theorem 4. *The endowment bias* $\kappa[v;\mu] = 0$ *if and only if* μ *is the uniform distribution on* 2^N*, the benchmark distribution for the Banzhaf value.*

Theorem 5. *Given* $\mu \in \mathcal{F}$*, the total bias* $\sum_i \kappa_i[v;\mu] = 0$ *if and only if* μ *is the uniform distribution on* 2^N*.*

The Shapley value, however, tends to demonstrate strong evidence of negative endowment bias, largely due to the diminishing marginal effect of v. This counterintuitive issue could bring the users undesirable inference from the data and eventually limits the usage of the value. When v is a set of modeling criteria, in general, the candidate variables are modestly correlated with each other; as a consequence, for any $i \notin T$, its explanatory and predictive power is partially mitigated by the members of T. Moreover, the larger the set T, the likely more the mitigation; thus, superaddtivity assumption is highly artificial, and diminishing marginality is more pervasive.

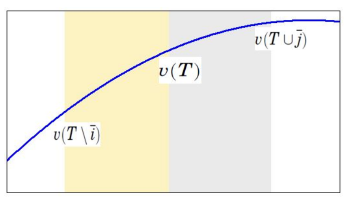

Figure 1: Diminishing Marginal Effect for a Typical Set T

Let us formally define the diminishing marginality. Ideally, we expect the inequality

$$v(T) - v(T \setminus \bar{i}) \geq v(T \cup \bar{j}) - v(T)$$

holds for a typical T, $i \in T$ and $j \notin T$ as in Figure 1. But it is laborious to locate the typical T and it is impractical for the inequality to hold for all T's. Thus, we average both sides of the inequality for all T's of size t, all $i \in T$ and all $j \notin T$. Obviously, the averages are

$$\omega_t \stackrel{\text{def}}{=} \frac{(t-1)!(n-t)!}{n!} \sum_{|T|=t} \sum_{i \in T} [v(T) - v(T \setminus \bar{i})], \quad t = 1, 2, \cdots, n;$$
$$\pi_t \stackrel{\text{def}}{=} \frac{t!(n-t-1)!}{n!} \sum_{|T|=t} \sum_{j \notin T} [v(T \cup \bar{j}) - v(T)], \quad t = 0, 1, \cdots, n-1.$$

We say v has *diminishing marginal effect* if $\omega_t \geq \pi_t$ for $t = 1, 2, \cdots, n-1$; we also say v has *diminishing marginal gain* (or *diminishing marginal loss*) if ω_t (or π_t, respectively) is a decreasing function of t.

Theorem 6. $\pi_t = \omega_{t+1}$ *for* $t = 0, 1, \cdots, n-1$. *Consequently, the following statements are equivalent: 1. v has diminishing marginal effect; 2. v has diminishing marginal gain; 3. v has diminishing marginal loss.*

Theorem 7. *Assume* $P_T = \frac{(|T|)!(n-|T|)!}{(n+1)!}$, *the benchmark prior for the Shapley value. If v has diminishing marginal effect, then* $\sum_{i=1}^{n} \kappa_i[v;\mu] \leq 0$; *if v is super-additive, then* $\sum_{i=1}^{n} \kappa_i[v;\mu] \geq 0$.

In an objective valuation, such as variable selection, either positive or negative endowment bias should be avoided. In (1), we place the same weight on the marginal loss as on the marginal gain. To mitigate the endowment bias, one could study the *unbiased D-value* defined as

$$\tilde{\psi}_i[v;\mu] \stackrel{\text{def}}{=} (1-\alpha)\gamma_i[v;\mu] + (1+\alpha)\lambda_i[v;\mu]. \tag{6}$$

where $\alpha = \frac{\sum_j \kappa_j[v;\mu]}{\sum_j \psi_j[v;\mu]}$, called *endowment bias ratio*, is the ratio between the total endowment bias and the total D-value. Clearly, there is no more endowment bias in the unbiased D-value:

$$\begin{aligned} & \sum_i (1-\alpha)\gamma_i[v;\mu] - \sum_i (1+\alpha)\lambda_i[v;\mu] \\ = \; & \sum_i (\gamma_i[v;\mu] - \lambda_i[v;\mu]) - \alpha \sum_i (\gamma_i[v;\mu] + \lambda_i[v;\mu]) = 0. \end{aligned}$$

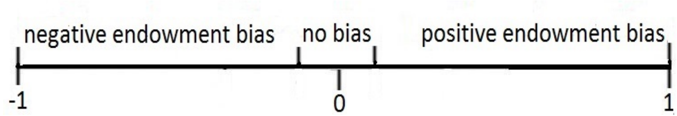

Figure 2: The Endowment Bias Ratio

In general, the endowment bias ratio α lies between -1 and 1 when v is a monotonic function. If α is close to 0, then there is no endowment bias. If

α is significantly positive, then $\sum_j \gamma_j[v;\mu] > \sum_j \lambda_j[v;\mu]$ and we have positive endowment bias; we should place more weights on the expected marginal loss. In contrast, if α is significantly negative, then we have negative endowment bias and we should place more weights on the expected marginal gain.

Theorem 8. *If $\mu \in \mathcal{F}$ and the size $|\mathbf{S}|$ has a distribution of Binomial(n,p), then the endowment bias ratio $\alpha = 2p-1$ and the unbiased D-value is*

$$\begin{aligned}\tilde{\psi}_i[v;\mu] &= 4(1-p)\gamma_i[v;\mu] = 4p\lambda_i[v;\mu] = 4p(1-p)\psi_i[v;\mu] \\ &= 4\sum_T p^{|T|}(1-p)^{n-|T|+1}\big[v(T) - v(T\setminus \bar{i})\big].\end{aligned}$$

If p itself is random and has a Beta(s,t) distribution, we marginalize out p in (5); the result is the expected D-value:

$$\begin{aligned}\mathrm{E}\psi_i[v;\mu] &= \sum_T \big[v(T) - v(T\setminus \bar{i})\big] \tfrac{1}{\beta(s,t)} \int_0^1 p^{|T|-1+s-1}(1-p)^{n-|T|+t-1}dp \\ &= \sum_T \tfrac{\beta(s+|T|-1,t+n-|T|)}{\beta(s,t)}\big[v(T) - v(T\setminus \bar{i})\big].\end{aligned} \tag{7}$$

In this generalization, $\beta(\cdot,\cdot)$ is the beta function and the model $\mathbf{S}$ has an expected size of $n\mathrm{E}[p] = \frac{ns}{s+t}$. Likewisely, by Theorem 8, the expected unbiased D-value is

$$\begin{aligned}\mathrm{E}\tilde{\psi}_i[v;\mu] &= 4\sum_T [v(T) - v(T\setminus \bar{i})] \tfrac{1}{\beta(s,t)} \int_0^1 p^{|T|+s-1}(1-p)^{n-|T|+1+t-1}dp \\ &= 4\sum_T \tfrac{\beta(s+|T|,t+n-|T|+1)}{\beta(s,t)}\big[v(T) - v(T\setminus \bar{i})\big].\end{aligned} \tag{8}$$

As a special case, when $s = t = 1$ (i.e. p has the uniform distribution on $[0,1]$), then (7) reduces to the Shapley value $\Psi_i[v]$ and consequently, (8) becomes its unbiased one as stated in Theorem 9.

Theorem 9. *The unbiased Shapley value is given by*

$$\tilde{\Psi}_i[v] = 4\sum_T \frac{(|T|)!(n-|T|+1)!}{(n+2)!}\big[v(T) - v(T\setminus \bar{i})\big].$$

In a subjective valuation, such as evaluating a used car, we need to consider a second layer of risk: the uncertainty in $v(T\cup\bar{i}) - v(T)$ and $v(T) - v(T\setminus\bar{i})$ differs and the former one is far more significant than the latter one. Consequently, a positive α is more likely for a risk-averse evaluator. This partially explains the vast existence of positive endowment bias.

4. Estimation

For a large n, exact calculation of the D-value $\psi[v;\mu]$ is not practical; thus we seek random sampling techniques to approximate it. An easy way to approximate the D-value $\psi[v;\mu]$ and its components by random sampling is to randomly draw many $\mathbf{S}$'s and then apply the definition of D-value. For a large n, however, some members in N could be much less represented in these $\mathbf{S}$'s than other members. In this section we study instead a random ordering in which each member appears exactly once and then calculate added value for all members in the ordering. Finally we average the added values from a large number of random orderings to estimate $\psi[v;\mu]$ and its components.

Let Ω be the set of orderings of all candidate variables. There are $n!$ orderings in total. We randomly take an ordering τ from Ω:

$$\tau : \quad \emptyset \to i_1 \to \cdots \to i \to \cdots \to i_n.$$

Let Ξ_i^τ be the set of variables in N which precede i in the ordering τ, and let

$$\phi_i^\tau = v(\Xi_i^\tau \cup \bar{i}) - v(\Xi_i^\tau).$$

Shapley(1953) showed that $\mathrm{E}[\phi_i^\tau] = \Psi_i[v]$ where the expectation is taken under the premise that any ordering is equally likely to be picked from Ω. To estimate $\gamma[v;\mu]$, $\lambda[v;\mu]$, and $\psi[v;\mu]$ for any μ, we bind the probability density to the sequential increments by letting

$$\tilde{\phi}_i^\tau \overset{\mathrm{def}}{=} \frac{(P_{\Xi_i^\tau} + P_{\Xi_i^\tau \cup \bar{i}})n!}{(|\Xi_i^\tau|)!(n - |\Xi_i^\tau| - 1)!} \left[v(\Xi_i^\tau \cup \bar{i}) - v(\Xi_i^\tau) \right].$$

Theorem 10. $\mathrm{E}[\tilde{\phi}_i^\tau] = \psi_i[v;\mu]$ *where* τ *has a uniform distribution on* Ω*. Furthermore,*

$$\begin{aligned} \gamma_i[v;\mu] &= \mathrm{E}\left[\frac{n!P_{\Xi_i^\tau \cup \bar{i}}}{(|\Xi_i^\tau|)!(n-|\Xi_i^\tau|-1)!}\big(v(\Xi_i^\tau \cup \bar{i}) - v(\Xi_i^\tau)\big)\right], \\ \lambda_i[v;\mu] &= \mathrm{E}\left[\frac{n!P_{\Xi_i^\tau}}{(|\Xi_i^\tau|)!(n-|\Xi_i^\tau|-1)!}\big(v(\Xi_i^\tau \cup \bar{i}) - v(\Xi_i^\tau)\big)\right]. \end{aligned} \tag{9}$$

In particular, the Shapley value can be decomposed into two parts

$$\begin{aligned} \gamma_i[v;\mu] &= \mathrm{E}\left[\tfrac{|\Xi_i^\tau|+1}{n+1}\big(v(\Xi_i^\tau \cup \bar{i}) - v(\Xi_i^\tau)\big)\right], \\ \lambda_i[v;\mu] &= \mathrm{E}\left[\tfrac{n-|\Xi_i^\tau|}{n+1}\big(v(\Xi_i^\tau \cup \bar{i}) - v(\Xi_i^\tau)\big)\right]. \end{aligned} \tag{10}$$

and the unbiased Shapley value equals

$$\tilde{\Psi}_i[v] = 4\mathrm{E}\left[\frac{(|\Xi_i^\tau|+1)(n-|\Xi_i^\tau|)}{(n+1)(n+2)}\big(v(\Xi_i^\tau \cup \bar{i}) - v(\Xi_i^\tau)\big)\right]. \tag{11}$$

To estimate these values, we take a large sample of orderings from Ω. Then we use the average of ϕ_i's in the sample to estimate $\Psi_i[v]$, use the average of $\tilde{\phi}_i$'s in the sample to estimate $\psi_i[v;\mu]$, and use (9) and (10) to estimate expected marginal gain and loss. The averages converge as the sample size increases, according to the large sample theories. The sampling error reduces as the sample size increases; its size is asymptotically approximated by the Central Limit Theorem. Additionally, we can extract the medians, confidence intervals, and other robust statistics from the large sample of sequential marginal gain.

The sequential approach implied by Theorem 10 is different from the classical stepwise procedures in regression. In the sequential approach, we average the changed v from directly nested models. The stepwise procedure admits and drops variables based on their significance test; however, the exact significance level cannot be calculated (cf Freedom, 1983). As a matter of fact, because of the diminishing marginality, variable i's significance tends to become smaller as the size of Ξ_i^τ increases; consequently, different procedures or starting from different models could lead to different selected models; thus stepwise procedures are sub-optimal. Another drawback of this procedure is that it heavily relies on a single criterion, such as the F-statistic.

When the model size $|\mathbf{S}|$ has a distribution of Binomial(n, p), the estimated model $\hat{\mathbf{S}}$ certainly depends on the choice of p; meanwhile, we can also estimate p from an estimated model $\hat{\mathbf{S}}$ using the relation $\mathrm{E}[|\mathbf{S}|] = np$. Thus, we could estimate both $\mathbf{S}$ and p recursively and iteratively by the following algorithm:

Step 1: use a non-informative prior, such as the benchmark distribution for the Shapley value or the Banzhaf value to estimate a $\hat{\mathbf{S}}$;

Step 2: estimate p using $\hat{p} = \frac{|\hat{\mathbf{S}}|}{n}$;

Step 3: use Binomial$(n, \hat{p})$ as the prior to estimate a new $\hat{\mathbf{S}}$;

Step 4: repeat Step 2 and Step 3 until $\hat{p}$ converges.

If the algorithm does not converge, then an extended algorithm called MCMC (Markov chain Monte Carlo) can be used to estimate the posterior distribution of $(p, \mathbf{S})$.

5. Empirical Studies

Our approach opens new areas of applications and suggests improvement on how to apply the Shapley value. In this section, we first attempt to solve a variable selection problem using a multi-criteria decision analysis based on unbiased multivariate Shapley value. We then extend the above binary categorization to ternary to calculate the effect of one or more schooling years on the hourly wage.

*5.1. **Unbiased Multivariate Shapley Value Analysis of Model Selection***

In this empirical study, we analyze the employment-related data in Arellano and Bond (1991) for 140 U.K. firms from 1976 to 1984. The objective is to model the employment size using a set of possible explanatory variables, including up to 2 lags of both the explanatory variables and the dependent variable. The variables are:

$E_{i,t}$: log of employment in company i at the end of year t,

$W_{i,t}$: log of real product wage in company i at the end of year t,

$K_{i,t}$: log of gross capital in company i at the end of year t,

$Y_{i,t}$: log of industry output in company i at the end of year t.

In this example, $E_{i,t}$ is the dependent variable; there are 11 candidate explanatory variables: $E_{i,t-1}$, $E_{i,t-2}$, $W_{i,t}$, $W_{i,t-1}$, $W_{i,t-2}$, $K_{i,t}$, $K_{i,t-1}$, $K_{i,t-2}$, $Y_{i,t}$, $Y_{i,t-1}$, and $Y_{i,t-2}$. Thus, there are totally $2^{11} = 2048$ potential models. Let all models also contain a common intercept, a time effect that is common to all companies, a permanent but unobservable firm-specific effect, and an error term. The models are estimated by the GMM method with all candidate variables and the constant intercept as the instruments.

Let the measurement function $v(T)$ be 5-dimensional. The first component is the goodness of fit, measured by the R-squared when modeling $E_{i,t}$ by the variables in T. The second and third components are its significance when a new variable is admitted to $\mathbf{S}$, using its squared t-statistic and absolute value of t-statistic, respectively. The fourth and fifth components are the predictive measure, using the mean squared error when the model of T is used for dynamic and static forecasts, respectively. Alternative criteria, such as the log likelihood, F-statistic, Theil-U2, etc, can also be used.

From the 11! total orderings of the independent variables, we randomly sample $5,000$. For each ordering τ, we run stepwise regressions from the emptyset $\emptyset$ to the grand coalition N to get the scaled marginal gain and scaled marginal loss in the ordering, using (10). Averaging these $5,000$ scaled marginal gains and scaled marginal losses, we obtain the estimated $\hat{\gamma}_i[v;\mu]$ and $\hat{\lambda}_i[v;\mu]$. Finally, we estimate the Shapley value by $\hat{\Psi}_i[v] = \hat{\gamma}_i[v;\mu] + \hat{\lambda}_i[v;\mu]$ and the unbiased Shapley value by formula (6). In Tables 2 and 3, we report both the original and the unbiased multivariate Shapley value, as well as its percentage share in the parenthesis. Estimated endowment bias ratio $\hat{\alpha}$ is in the last row of Table 3.

Based on these two tables, we definitely need to keep the variables $E_{i,t-1}$, $K_{i,t}$, and $K_{i,t-1}$. They perform significantly well under all 5 criteria. Five other variables, $E_{i,t-2}$, $W_{i,t}$, $W_{i,t-2}$, $K_{i,t-2}$, and $Y_{i,t-2}$, perform well under some, but not all criteria. For parsimony purpose, we should drop $W_{i,t-1}$, $Y_{i,t}$, and $Y_{i,t-1}$. Thus, the final model consists of the explanatory variables $E_{i,t-1}$, $K_{i,t}$, $K_{i,t-1}$, $E_{i,t-2}$, $W_{i,t}$, $W_{i,t-2}$, $K_{i,t-2}$, and $Y_{i,t-2}$. Mathematically speaking, the decision rule is to drop a variable if its percentage shares are less than 5% in all 5 criteria.

The estimation results in Table 3 shows strong evidence of negative endowment bias for the model fit criterion, and even larger negative endowment bias for the predictive criteria. We also find that the employment rigidity, indicated by $E_{i,t-1}$, is the most important factor in determining the employment size. In addition, one could reasonably argue that the employment size largely depends on the gross capital (both prevailing and 1-year lagged), and barely relies on the prevailing wage and the 2-year lagged industrial output.

5.2. ***Worth of Additional Schooling Years***

In this study, we extend the idea of dichotomy to trichotomize the marginal effect and answer a question extensively studied: by how much would a higher level of education likely raise one's hourly pay rate. There are of course many other factors which also affect the hourly income. To remove the effect of a worker's natural ability, his family background, and his innate intelligence on income, we analyze the Twinsburg schooling data in Ashenfelter and Krueger(1994) and Ashenfelter and Rouse (1998).

The data contains several income-related variables for 340 pairs of identical twins. For simplicity, we only assume that marriage status and union coverage, besides educational level, also affect one's income. For each pair of twins, we randomly name them "Type 1" and "Type 2"; thus, we create a randomized dataset. For any randomized dataset, let us categorize the ith pair of twins by introducing the union coverage variable U, the marriage status variable M, and the educational level variable E, defined as

$$U_i = \begin{cases} -1, & \text{if Type 2 has union coverage but Type 1 hasn't;} \\ 1, & \text{if Type 1 has union coverage but Type 2 hasn't;} \\ 0, & \text{otherwise.} \end{cases}$$

$$M_i = \begin{cases} -1, & \text{if Type 2 is married but Type 1 isn't;} \\ 1, & \text{if Type 1 is married but Type 2 isn't;} \\ 0, & \text{otherwise.} \end{cases}$$

$$E_i = \begin{cases} -1, & \text{if Type 2 has one or more additional schooling years than Type 1;} \\ 1, & \text{if Type 1 has one or more additional schooling years than Type 2;} \\ 0, & \text{otherwise.} \end{cases}$$

Accordingly, we classify the 340 observations into 27 categories, some of which may contain no observations. For Table 4, the data is actually not randomized: we label Type 1 twins before Type 2 twins according to their orders in the original data. Column 2 lists the frequency for each category of (U, M, E) in Column 1.

Our purpose is to quantify the income differential due to the schooling differential only, U and M remaining unchanged. Let $v(U, M, E)$ be the average difference of hourly wages between Type 1 twins and Type 2 twins in the category (U, M, E). The difference comes from the values of U, M, E, as well as sampling error. Column 3 of Table 4 lists the v for the non-randomized dataset. For the category $(U, M, 1)$, the marginal gain and effect is $v(U, M, 1) - v(U, M, 0)$; the effect is only due to the difference in E_i, plus a random error. For the category $(U, M, -1)$, the marginal loss and effect is $v(U, M, 0) - v(U, M, -1)$. For the category $(U, M, 0)$, the marginal gain is $v(U, M, 0) - v(U, M, -1)$ and the marginal loss is $v(U, M, 1) - v(U, M, 0)$; we average them for calculating the marginal effect, i.e.

$$\frac{[v(U, M, 0) - v(U, M, -1)] + [v(U, M, 1) - v(U, M, 0)]}{2} = \frac{v(U, M, 1) - v(U, M, -1)}{2}.$$

The last three columns of Table 4 list the marginal gain, loss, and effect for the non-randomized dataset. The last row is the weighted marginal gain, loss and effect weighted by the frequencies. For example, the weighted marginal effect is calculated as:

$$\Big[1(-11.791) + 2(-3.4492) + \cdots + 2(4.7643) + 1(-.2133)\Big]/338 = 2.7995.$$

Repeating the above calculations for a large number of randomized datasets, we obtain a large set of frequency-weighted marginal gain, loss and effect, due to one or more years of schooling. Given the average hourly wage \$14.44 in 1993, we find that both mean and median wage increase is about 18.3%. The results are consistent with, though slightly lower than, those from the marginal gain and marginal loss. The statistics are summarized in Table 1.

Table 1: Marginal Gain, Loss and Effect of 1+ Years of Schooling (in %, except α)

	Mean	Median	Std. Dev	.95 Confidence Band
marginal effect	18.2879	18.2833	1.30667	[15.8921, 20.7155]
marginal gain	18.8138	18.7626	4.30053	[10.5990, 27.2509]
marginal loss	18.7867	18.7543	4.28981	[10.5758, 27.1039]
bias ratio α	.001343	.000459	.448105	[-.864527, .867927]

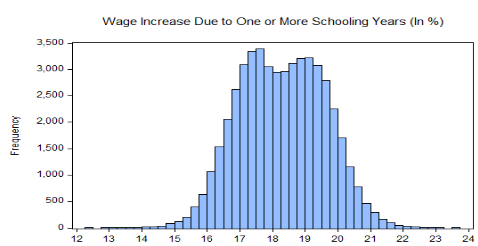

Figure 3: Multi-modal Effects of Additional Schooling Years on Wage

Though the estimated endowment bias ratio has slightly positive mean and median, it is not significantly different from zero, showing no enough evidence of endowment bias. Interestingly, the effects show a multi-modal distribution (Figure 3) with modes at 17.7% and 19.2%, respectively. Unlike a regression method, our calculation allows heterogeneous effects from the other explanatory variables − the heterogeneity significantly presents in Table 4. In addition, our approach addresses the asymmetric sides of the marginal effects and the non-parametric method assumes less assumptions than a regression one. Interested readers can re-categorize the E_i variable to find the effects of other educational differential.

6. Discussions, Variations, and Applications

Given the evaluation $\tilde{\psi}$, we select those variables with large D-value $\tilde{\psi}_i$ in modeling the data generating process. The selected variables, however, may not have the best collective performance under certain criterion. The model with the best performance under the criterion, on the other hand, does not necessarily produce the best performance under other criteria. This dilemma leads us to evaluate the variables not in one specific model context, but in modeling scenarios specified by a prior distribution. Moreover, as a tool for multi-criteria decision analysis, our valuation approach seeks the balance between data fitting, prediction, and other criteria. Depending on specific contexts, similar ideas can be applied to other areas of economics, finance, political science, and statistics without substantial changes.

6.1. Structural valuation

In a real valuation situation, it could be proper to specify a context-specific prior μ; for example, in a voting, let $|\mathbf{S}|$ have a uniform distribution on the integers in $\lfloor n/4, 3n/4 \rfloor$, rather than on the integers in $[0, n]$. This can also be readily done by placing restrictions on μ; for example, if players i and j would never cooperate in the voting, then the probability of $\overline{ij} \in \mathbf{S}$ should be zero. Though analytic formula is unlikely available for $\psi[v; \mu]$ under a restricted μ, Monte Carlo method is generally feasible. In the Monte Carlo simulations, we simply ignore the restricted cases, randomly generated from Ω or a $\mu \in \mathcal{F}$.

6.2. Unemployment Compensation

Let N be the labor force (either employed or seeking for employment) of an economy, $\mathbf{S}$ be the employed, and $v(T)$ be the social welfare or surplus made collectively by T. Then, a fair and efficient rule to distribute unemployment benefits could be the solution of $\psi[v;\mu]$ such that $\mu \in \mathcal{F}$ and $\sum_{i=1}^{n} \psi_i[v;\mu] = \max_T v(T)$, instead of $v(N)$. Under the efficiency assumption of the labor market, the employed $\mathbf{S}$ is expected to maximize the social welfare of the nation; thus the distribution rule solves $\sum_{i=1}^{n} \psi_i[v;\mu] = \mathrm{E}[v(\mathbf{S})]$.

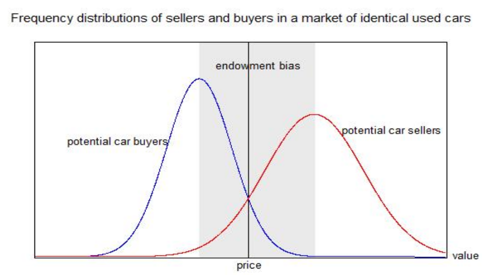

Figure 4: The price is determined at the intersection of frequency plots

6.3. A Theory of Price

Consider a market of identical used cars (or a secondary market of a stock or bond). There are two groups of market participants: potential sellers and potential buyers. Each group has its own frequency distribution with heterogeneous valuation about the car. The value of a car to the potential sellers is the marginal gain; to the potential buyers, it is the marginal loss. In Figure 4, the frequence plots intersect at a price where the demand quantity equals to the supply quantity.

6.4. Power Index in a Ternary Voting System

Let us study a ternary voting rule in which a voter has only two choices, but $v(\mathbf{S})$ has three outcomes: passed, failed, or dropped. Here $\mathbf{S}$ denotes the random coalition of voters who vote for the bill.

Let τ be a random ordering of all voters. In this ordering, there is exactly one pivotal voter i such that he changes the outcome $v(\Xi_i^\tau)$, but $N \setminus \Xi_i^\tau \setminus \bar{i}$ no longer changes the outcome $v(\Xi_i^\tau \cup \bar{i})$. The chance of being pivotal in all orderings can be used to measure one's power in the voting system.

6.5. *Value of a Coalition Z*

If N is a firm, a coalition Z could be the sales group or the R&D team. We treat coalition Z as an indivisible entity; and there are many value-added opportunities for Z and the rest of the firm. For a value of Z, a natural extension to the D-value ψ_i is

$$\mathrm{E}\left[v(\mathbf{S} \cup Z) - v(\mathbf{S} \setminus Z)\right]$$

6.6. *Value of Integration or M&A*

Assume two disjoint firms M and N. We consider the potential integration between M and N. To evaluate the integration, one could work from the added value between their coalitions:

$$\mathrm{E}[v(\mathbf{X} \cup \mathbf{Y}) - v(\mathbf{X}) - v(\mathbf{Y})]$$

where random $\mathbf{X} \subseteq N$ and $\mathbf{Y} \subseteq M$ are potential cooperating coalitions.

6.7. *Value of Substitution*

If $i \in \mathbf{S}$, then i's position could be replaced by some other player, say j, in $N \setminus \mathbf{S}$. So we have the added value

$$v(\mathbf{S} \cup \bar{j} \setminus \bar{i}) - v(\mathbf{S}).$$

Dually, if $i \notin \mathbf{S}$, then i could substitute for a player, say j, in $\mathbf{S}$. This substitution gives an added value:

$$v(\mathbf{S} \cup \bar{i} \setminus \bar{j}) - v(\mathbf{S}).$$

6.8. *Analysis of Variance*

In regressing Y on the variables in T by a linear model, we let $v(T)$ be the variance of Y explained by T. Decomposition of $v(N) - v(\emptyset)$ by the Shapley value provides a specific share each variable contributes to modeling Y using all variables in N.

6.9. Outliers Detection

In detecting outliers observations, one could apply the D-value $\tilde{\psi}_i$ to evaluate the overall performance of each observation. For example, consider the MCD estimator of robust covariance (cf Rousseeuw, 1985). Among all covariances of subsamples of a given size, the MCD estimator has the minimum determinant. We let N be the full sample of observations and T be a subsample. Let $v(T)$ be the determinant of the covariance of the subsample T. We can reasonably assume no discrimination on any subsample among all subsamples of the same size. A binomial(n, p) prior with $p \in [.9, .99]$ could be a good choice for the size of non-outliers observations.

6.10. Continuous Categorization

In the empirical study of schooling years, we replace the binary classification with an ordered ternary one. In a more generic situation, we may assume a continuous random vector $\mathbf{X} = (X_1, \cdots, X_n)' \in R^n$ and a differentiable function $v : R^n \to R$. Let μ be the distribution of $\mathbf{X}$. In this setting, the marginal effect is represented by a partial derivative and the corresponding valuation of variable X_i is

$$\psi_i[v; \mu] = \mathrm{E}\left[\frac{dv(\mathbf{X})}{dx_i}\right] = \int_{(x_1,\cdots,x_n)\in R^n} \frac{dv(x_1, \cdots, x_n)}{dx_i} d\mu \tag{12}$$

in contrast with (2) and the Aumann-Shapley value (Aumann and Shapley, 1974). All interdependence, restrictions, and uncertainty assumption about $\mathbf{X}$ are contained in the distribution μ.

Given a sample of $X_1, \cdots, X_n$ and the value v at the sample observations, we may use a multivariate empirical distribution $\hat{\mu}$ to fit the distribution μ and apply a multivariate differentiable function $\hat{v}$ to fit v. Finally, formula (12) with $\mu = \hat{\mu}$ and $v = \hat{v}$ estimates variable X_i's relative importance in modeling Y with a differentiable function v.

References

[1] M. Arellano, S. Bond, *Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations*, Rev. Econ. Stud. 58 (1991), 277-297.

[2] O. Ashenfelter, A. Krueger, *Estimates of the economic return to Schooling from a new sample of twins*, Amer. Econ. Rev. 84(1994), 1157-1173.

[3] O. Ashenfelter, C. Rouse, *Income, schooling, and ability: evidence from a new sample of identical twins*, Quart. J. Econ. 113(1998), 253-284.

[4] R.J. Aumann, L.S. Shapley, *Values of Non-Atomic Games*. Princeton University Press, Princeton, NJ (1974).

[5] J.F. Banzhaf, *Weighted voting doesn't work: a mathematical analysis*, Rutgers Law Rev. 19(1965), 317-343.

[6] G.E.P. Box, *Science and Statistics*, J. of Amer. Stat. Assoc. 71(1976), 791-799.

[7] N. Budina, B. Gracia, X. Hu, S. Saksonovs, *Recognizing the Bias: Financial Cycles and Fiscal Policy*, IMF Working Paper No. 15/246 (2015).

[8] M. Clyde, E. I. George, *Model Uncertainty*, Statistical Science 19(2004), 81-94.

[9] F. Devicienti, *Shapley-value decomposition of changes in wage distributions: a note*, J. Econ. Inequal 8(2010), 35-45.

[10] D.A. Freedman, *A Note on Screening Regression Equations*, Amer. Statistician 37(1983), 152-155.

[11] E.I. George, R.E. McCulloch, *Approaches for Bayesian variable selection*, Statistica Sinica 7(1997), 339-373.

[12] U. Gromping, *Estimators of relative importance in linear regression based on variance decomposition*, Amer. Statistician 61(2007), 1-9.

[13] X. Hu, *Value of Loss in n-Person Games*, UCLA Comput. Appl. Math. Reports 02-53 (2002).

[14] X. Hu, *An asymmetric Shapley-Shubik power index*, Intl. J. Game Theory 34(2006), 229-240.

[15] O. Israeli, *A Shapley-based decomposition of the R-Square of a linear regression*, J. Econ. Inequal 5(2007), 199-212.

[16] S. Lipovetsky, M. Conklin, *Analysis of regression in game theory approach*, Appl. Stoch. Models in Business Industry 17(2001), 319-330.

[17] N. Megiddo, On Finding Additive, Superadditive and Subadditive Set-Function Subject to Linear Inequalities, RJ 6329 (61998) 7/11/88, Computer Science/Nathematics, IBM Almaden Research Center, San Jose, California.

[18] D. Monderer, D. Samet, *Variations on the Shapley Value*. In: R.J. Aumann and S. Hart (eds) Handbook of Game Theory, vol 2, Chapter 54 (2002), pp.2055-2076. Elsevier Science, Netherlands.

[19] R.B. O'Hara, M.J. Sillanpaa, *A Review of Bayesian Variable Selection Methods: What, How and Which*, Bayesian Analysis 4(2009), 85-118.

[20] P.J. Rousseeuw, *Multivariate estimation with high break-down point*. In: W. Grossmann et al. (eds.) Mathematical Statistics and Applications, vol. B(1985), pp.283-297. Akademiai Kiado, Budapest.

[21] L.S. Shapley, *A value for n-person games*. In: H.W. Kuhn and A.W. Tucker (eds.) Contributions to the Theory of Games, Annals of Math. Studies, vol. 28 (1953), pp.307-317. Princeton University Press, Princeton, New Jersey.

[22] E. Winter, *The Shapley Value*. In: R.J. Aumann and S. Hart (eds) Handbook of Game Theory, vol 2, Chapter 53 (2002), pp.2025-2054. Elsevier Science, Netherlands.

Appendix

For notational simplicity, we assume that v is 1-dimensional set function $v : 2^N \to R$ in the proofs.

A1. Proof of Theorem 1

For $t = 0, 1, \cdots, n$, let $\delta_t \overset{\text{def}}{=} \sum_{|T|=t} P_T$, the probability of $|\mathbf{S}| = t$. As $\mu \in \mathcal{F}$, $P_T = \frac{(|T|)!(n-|T|)!}{n!}\delta_{|T|}$. For any fixed $i \in N$, we write (1) as

$$\begin{aligned}
\psi_i[v;\mu] &= \sum_{T \ni i} P_T[v(T) - v(T \setminus \bar{i})] + \sum_{Z \not\ni i} P_Z[v(Z \cup \bar{i}) - v(Z)] \\
&\overset{Q=T\setminus\bar{i}}{=} \sum_{T \ni i} P_T v(T) - \sum_{Q \not\ni i} P_{Q\cup\bar{i}} v(Q) + \sum_{Z \not\ni i} P_Z[v(Z \cup \bar{i}) - v(Z)] \\
&= \sum_{T \ni i} P_T v(T) + \sum_{Z \not\ni i} P_Z v(Z \cup \bar{i}) - \sum_{Z \not\ni i} [P_Z + P_{Z\cup\bar{i}}] v(Z) \qquad (13) \\
&\overset{T=Z\cup\bar{i}}{=} \sum_{T \ni i} P_T v(T) + \sum_{T \ni i} P_{T\setminus\bar{i}} v(T) - \sum_{Z \not\ni i} [P_Z + P_{Z\cup\bar{i}}] v(Z) \\
&= \sum_{T \ni i} v(T)[P_T + P_{T\setminus\bar{i}}] - \sum_{T \not\ni i} v(T)[P_T + P_{T\cup\bar{i}}].
\end{aligned}$$

Therefore, By (3)

$$\begin{aligned}
v(N) - v(\emptyset) &= \sum_{i \in N} \psi_i[v;\mu] \\
&= \sum_{i \in N} \sum_{T \ni i} v(T)[P_T + P_{T\setminus\bar{i}}] - \sum_{i \in N} \sum_{T \not\ni i} v(T)[P_T + P_{T\cup\bar{i}}] \\
&= \sum_{T \subseteq N} v(T) \sum_{i \in T} [P_T + P_{T\setminus\bar{i}}] - \sum_{T \subseteq N} v(T) \sum_{i \notin T} [P_T + P_{T\cup\bar{i}}] \\
&= \sum_{T \subseteq N} v(T) \left[|T| P_T + \sum_{i \in T} P_{T\setminus\bar{i}} - (n - |T|) P_T - \sum_{i \notin T} P_{T\cup\bar{i}} \right] \\
&= \sum_{T \subseteq N} v(T) \left[(2|T| - n) P_T + \sum_{i \in T} P_{T\setminus\bar{i}} - \sum_{i \notin T} P_{T\cup\bar{i}} \right].
\end{aligned} \qquad (14)$$

We compare the coefficients of $v(N)$ and $v(\emptyset)$ in (14) to get

$$\begin{cases} n\delta_n + \delta_{n-1} = 1, \\ n\delta_0 + \delta_1 = 1. \end{cases} \qquad (15)$$

For any T such that $T \neq N$ and $T \neq \emptyset$, the coefficients of $v(T)$ in (14) imply that

$$(2|T| - n) P_T = \sum_{i \notin T} P_{T\cup\bar{i}} - \sum_{i \in T} P_{T\setminus\bar{i}}.$$

Let $|T| = t$, $1 \leq t \leq n-1$, and rewrite the above equation in terms of δ's,

$$(2t-n)\frac{t!(n-t)!}{n!}\delta_{t} = (n-t)\frac{(t+1)!(n-t-1)!}{n!}\delta_{t+1} - t\frac{(t-1)!(n-t+1)!}{n!}\delta_{t-1}.$$

Or simply,

$$(2t-n)\delta_t = (t+1)\delta_{t+1} - (n-t+1)\delta_{t-1}. \tag{16}$$

By (15), (16) and $\sum_{t=0}^{n} \delta_t = 1$, we have $n+2$ linear equations of $n+1$ unknowns,

$$\begin{bmatrix} n & 1 & & & & & & & \\ n & 2-n & -2 & & & & & & \\ & n-1 & 4-n & -3 & & & & & \\ & & & \ddots & \ddots & \ddots & & & \\ & & & & & 2 & n-2 & -n \\ & & & & & & 1 & n \\ 1 & 1 & 1 & \cdots & \cdots & 1 & 1 & 1 \end{bmatrix} \begin{pmatrix} \delta_0 \\ \delta_1 \\ \delta_2 \\ \vdots \\ \delta_{n-2} \\ \delta_{n-1} \\ \delta_n \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

The rank of the $(n+2) \times (n+1)$ coefficient matrix is at least n, due to the fact that the submatrix of the first n columns and the middle n rows has determinant $n! \neq 0$. It is not hard to verify that the general solution to the system of equations is

$$\delta_t = \frac{1}{n+1} + (-1)^t \binom{n}{t} \eta \tag{17}$$

for some indeterminate η. Any η with $|\eta| \leq \min_{0 \leq s \leq n} \frac{s!(n-s)!}{(n+1)!} = \frac{(\lfloor \frac{n+1}{2} \rfloor)!(n - \lfloor \frac{n+1}{2} \rfloor)!}{(n+1)!}$ guarantees non-negative δ_t's. Thus, for any $T \subseteq N$,

$$P_{T} = \frac{(|T|)!(n-|T|)!}{(n+1)!} + (-1)^{|T|}\eta.$$

Therefore, for any T with $i \in T$,

$$P_{T} + P_{T \setminus \bar{i}} = \frac{(|T|-1)!(n-|T|)!}{n!}, \tag{18}$$

and for any T with $i \notin T$,

$$P_{T} + P_{T \cup \bar{i}} = \frac{(|T|)!(n-|T|-1)!}{n!}. \tag{19}$$

Finally, we plug (18) and (19) into (13),

$$\begin{aligned}
\psi_i[v;\mu] &= \sum_{T\ni i}\tfrac{(|T|-1)!(n-|T|)!}{n!}v(T) - \sum_{T\not\ni i}\tfrac{(|T|)!(n-|T|-1)!}{n!}v(T) \\
&\overset{Z=T\cup\bar{i}}{\equiv} \sum_{T\ni i}\tfrac{(|T|-1)!(n-|T|)!}{n!}v(T) - \sum_{Z\ni i}\tfrac{(|Z|-1)!(n-|Z|)!}{n!}v(Z\setminus\bar{i}) \qquad (20)\\
&\overset{T=Z}{\equiv} \sum_{T\subseteq N}\tfrac{(|T|-1)!(n-|T|)!}{n!}[v(T)-v(T\setminus\bar{i})] = \Psi_i[v].
\end{aligned}$$

A2. Proof of Corollary 1

If $P_{_T} = \frac{(|T|)!(n-|T|)!}{(n+1)!} + (-1)^{|T|}\eta$, clearly $\mu \in \mathcal{F}$. Furthermore, we repeat (18)−(20) to get $\psi[v;\mu] = \Psi[v]$.

A3. Proof of Theorem 2

If $P_{_T} = \frac{1}{2^n} + (-1)^{|T|}\eta$ for any T, then $P_{_T} + P_{_{T\setminus\bar{i}}} = P_{_T} + P_{_{T\cup\bar{j}}} = \frac{1}{2^{n-1}}$ for any $i \in T$ and any $j \notin T$. By (13),

$$\begin{aligned}
\psi_i[v;\mu] &= \sum_{T\ni i}\tfrac{1}{2^{n-1}}v(T) - \sum_{T\not\ni i}\tfrac{1}{2^{n-1}}v(T) \\
&\overset{Z=T\cup\bar{i}}{\equiv} \tfrac{1}{2^{n-1}}\sum_{T\ni i}v(T) - \tfrac{1}{2^{n-1}}\sum_{Z\ni i}v(Z\setminus\bar{i}) \\
&\overset{Z=T}{\equiv} \tfrac{1}{2^{n-1}}\sum_{T\ni i}[v(T)-v(T\setminus\bar{i})] = b_i[v].
\end{aligned}$$

On the contrary, if $\psi[v;\mu] = b[v]$, then for any i,

$$\psi_i[v;\mu] = \frac{1}{2^{n-1}}\sum_{T}[v(T)-v(T\setminus\bar{i})] = \frac{1}{2^{n-1}}\sum_{T\ni i}v(T) - \frac{1}{2^{n-1}}\sum_{T\not\ni i}v(T).$$

By (13),

$$\begin{cases} P_{_T} + P_{_{T\setminus\bar{i}}} = \frac{1}{2^{n-1}}, & \text{for any } i \in T; \\ P_{_T} + P_{_{T\cup\bar{i}}} = \frac{1}{2^{n-1}}, & \text{for any } i \notin T. \end{cases} \qquad (21)$$

Without loss of generality, let $P_{_\emptyset} = \frac{1}{2^n} + \eta$ for some η with $|\eta| \le \frac{1}{2^n}$. In (21), let $T = \bar{i}$ and we get

$$P_{_T} = \frac{1}{2^{n-1}} - P_{\emptyset} = \frac{1}{2^n} + (-1)^{|T|}\eta \qquad (22)$$

for any T with $|T| = 1$. Let us assume the above identity holds for all T's with $|T| = t$ and consider any Z with $|Z| = t+1$. Pick an $i \in Z$ and apply (21),

$$P_{_Z} = \frac{1}{2^{n-1}} - P_{_{Z\setminus\bar{i}}} = \frac{1}{2^{n-1}} - \left[\frac{1}{2^n} + (-1)^t\eta\right] = \frac{1}{2^n} + (-1)^{|Z|}\eta.$$

This shows (22) is also true for Z with $|Z| = t+1$. By mathematical induction, we have proved (22) for all T's.

A4. Proof of Theorem 3

If $\mu \in \mathcal{F}$ and $|\mathbf{S}|$ has a binomial distribution with parameters n and p, then

$$P_{_T} = \frac{\delta_{_{|T|}}}{\begin{pmatrix} n \\ |T| \end{pmatrix}} = \frac{\begin{pmatrix} n \\ |T| \end{pmatrix} p^{|T|}(1-p)^{n-|T|}}{\begin{pmatrix} n \\ |T| \end{pmatrix}} = p^{|T|}(1-p)^{n-|T|}.$$

Next,

$$\begin{aligned}
\psi_i[v;\mu] &= \sum_{T \ni i} P_{_T}[v(T) - v(T \setminus \bar{i})] + \sum_{Z \not\ni i} P_{_Z}[v(Z \cup \bar{i}) - v(Z)] \\
&\overset{T=Z\cup\bar{i}}{=} \sum_{T \ni i} P_{_T}[v(T) - v(T \setminus \bar{i})] + \sum_{T \ni i} P_{_{T\setminus\bar{i}}}[v(T) - v(T \setminus \bar{i})] \\
&= \sum_{T \ni i} \left[p^{|T|}(1-p)^{n-|T|} + p^{|T|-1}(1-p)^{n-|T|+1}\right][v(T) - v(T \setminus \bar{i})] \\
&= \sum_{T \ni i} p^{|T|-1}(1-p)^{n-|T|}[v(T) - v(T \setminus \bar{i})].
\end{aligned}$$

From the above, we also see

$$\begin{aligned}
\gamma_i[v;\mu] &= \sum_{T \ni i} p^{|T|}(1-p)^{n-|T|}[v(T) - v(T \setminus \bar{i})] = p\psi_i[v;\mu], \\
\lambda_i[v;\mu] &= \sum_{T \ni i} p^{|T|-1}(1-p)^{n-|T|+1}[v(T) - v(T \setminus \bar{i})] = (1-p)\psi_i[v;\mu].
\end{aligned}$$

A5. Proof of Lemma 1

$$
\begin{aligned}
\kappa_i &= \mathrm{E}\left[v(\mathbf{S}) - v(\mathbf{S} \setminus \bar{i})\right] - \mathrm{E}\left[v(\mathbf{S} \cup \bar{i}) - v(\mathbf{S})\right] \\
&= \sum_{T \ni i} P_T \left[v(T) - v(T \setminus \bar{i})\right] - \sum_{Z \not\ni i} P_Z \left[v(Z \cup \bar{i}) - v(Z)\right] \\
&= \left[\sum_{T \ni i} P_T v(T) + \sum_{Z \not\ni i} P_Z v(Z)\right] - \sum_{T \ni i} P_T v(T \setminus \bar{i}) - \sum_{Z \not\ni i} P_Z v(Z \cup \bar{i}) \\
&\overset{Q = Z \cup \bar{i}}{=} \sum_{T} P_T v(T) - \sum_{T \ni i} P_T v(T \setminus \bar{i}) - \sum_{Q \ni i} P_{Q \setminus \bar{i}} v(Q) \\
&\overset{Z = T \setminus \bar{i}}{=} \sum_{T} P_T v(T) - \sum_{Z \not\ni i} P_{Z \cup \bar{i}} v(Z) - \sum_{Q} P_{Q \setminus \bar{i}} v(Q) + \sum_{Q \not\ni i} P_{Q \setminus \bar{i}} v(Q) \\
&= \sum_{T} \left[P_T - P_{T \setminus \bar{i}}\right] v(T) - \sum_{Z \not\ni i} P_{Z \cup \bar{i}} v(Z) + \sum_{Q \not\ni i} P_Q v(Q) \\
&= \sum_{T} \left[P_T - P_{T \setminus \bar{i}}\right] v(T) - \sum_{Z} P_{Z \cup \bar{i}} v(Z) + \sum_{Z \ni i} P_{Z \cup \bar{i}} v(Z) + \sum_{Q \not\ni i} P_Q v(Q) \\
&= \sum_{T} \left[P_T - P_{T \setminus \bar{i}} - P_{T \cup \bar{i}}\right] v(T) + \left[\sum_{Z \ni i} P_Z v(Z) + \sum_{Q \not\ni i} P_Q v(Q)\right] \\
&= \sum_{T} \left[P_T - P_{T \setminus \bar{i}} - P_{T \cup \bar{i}}\right] v(T) + \sum_{T} P_T v(T) \\
&= \sum_{T} \left[2P_T - P_{T \cup \bar{i}} - P_{T \setminus \bar{i}}\right] v(T).
\end{aligned}
$$

A6. Proof of Theorem 4

If $P_T = \frac{1}{2^n}$ for any T, then $2P_T - P_{T \cup \bar{i}} - P_{T \setminus \bar{i}} = 0$ for any i and T. By Lemma 1,

$$\kappa_i[v; \mu] = \sum_{T} \left[2P_T - P_{T \cup \bar{i}} - P_{T \setminus \bar{i}}\right] v(T) = \sum_{T} 0 * v(T) = 0.$$

On the contrary, if $\kappa[v; \mu] = 0$, then Lemma 1 implies that $2P_T - P_{T \cup \bar{i}} - P_{T \setminus \bar{i}} = 0$ for any i and T. We let the size of T run from 1 to n:

1. If $T = \bar{i}$, then $2P_{\bar{i}} - P_{\bar{i} \cup \bar{i}} - P_{\bar{i} \setminus \bar{i}} = 0$, showing $P_{\bar{i}} = P_{\emptyset}$ for any i.
2. If $T = \overline{ij}$ with $i \neq j$, then $2P_{\overline{ij}} - P_{\overline{ij} \cup \bar{i}} - P_{\overline{ij} \setminus \bar{i}} = 0$, showing $P_{\overline{ij}} = P_{\bar{j}} = P_{\emptyset}$ for any $i \neq j$.
3. Let us assume that $P_Z = P_{\emptyset}$ for any Z with $|Z| = t$. For any T with $|T| = t + 1$, we pick an i from T and apply the relation $2P_T - P_{T \cup \bar{i}} - P_{T \setminus \bar{i}} = 0$, showing $P_T = P_{T \setminus \bar{i}} = P_{\emptyset}$.

By mathematical induction, we conclude that $P_T = P_{\emptyset}$ for any T. As $\sum_{T} P_T = 1$, $P_T = \frac{1}{2^n}$.

A7. Proof of Theorem 5

For $\mu \in \mathcal{F}$, $P_{_T} = \frac{(|T|)!(n-|T|)!\delta_{_{|T|}}}{n!}$ where δ_t is the probability of $|\mathbf{S}| = t$. We first simplify the total bias for $\mu \in \mathcal{F}$.

$$\begin{aligned}
\sum_i \gamma_i[v;\mu] &= \sum_i \sum_{T \ni i} P_{_T}[v(T) - v(T \setminus \bar{i})] \\
&= \sum_{T \neq \emptyset} \frac{(|T|)!(n-|T|)!\delta_{_{|T|}}}{n!} \sum_{i \in T}[v(T) - v(T \setminus \bar{i})]. \\
\sum_i \lambda_i[v;\mu] &= \sum_i \sum_{T \not\ni i} P_{_T}[v(T \cup \bar{i}) - v(T)] \\
&\overset{Z=T\cup\bar{i}}{=} \sum_i \sum_{Z \ni i} P_{_{Z\setminus\bar{i}}}[v(Z) - v(Z \setminus \bar{i})] \\
&= \sum_{Z \neq \emptyset} \frac{(|Z|-1)!(n-|Z|+1)!\delta_{_{|Z|-1}}}{n!} \sum_{i \in Z}[v(Z) - v(Z \setminus \bar{i})].
\end{aligned}$$

Then,

$$\sum_i \kappa_i = \sum_{T \neq \emptyset} \frac{(|T|-1)!(n-|T|)!(|T|\delta_{_{|T|}} - (n-|T|+1)\delta_{_{|T|-1}})}{n!} \sum_{i \in T}\left[v(T) - v(T \setminus \bar{i})\right]. \tag{23}$$

Thus, $\sum\limits_i \kappa_i[v;\mu] = 0$ if and only if $|T|\delta_{_{|T|}} = (n-|T|+1)\delta_{_{|T|-1}}$ for any $T \neq \emptyset$. By induction on the size of T from 1 to n, we have

$$\begin{aligned}
\delta_1 &= n\delta_0, \\
\delta_2 &= \tfrac{n-1}{2}\delta_1 = \tfrac{n(n-1)}{2!}\delta_0, \\
\delta_3 &= \tfrac{n-2}{3}\delta_2 = \tfrac{n(n-1)(n-2)}{3!}\delta_0, \\
& \quad \cdots\cdots
\end{aligned}$$

which establishes that $\delta_t = \frac{n!}{t!(n-t)!}\delta_0$ for any $t \geq 1$. Finally, as $\sum\limits_{t=0}^{n} \delta_t = 1$, we get $\delta_0 = \frac{1}{2^n}$; therefore

$$P_{_T} = \frac{(|T|)!(n-|T|)!}{n!}\delta_{_{|T|}} = \frac{(|T|)!(n-|T|)!}{n!}\frac{n!}{(|T|)!(n-|T|)!}\delta_0 = \frac{1}{2^n}.$$

A8. Proof of Theorem 6

By definitions,

$$\begin{aligned}
\pi_t &= \tfrac{t!(n-t-1)!}{n!} \sum_{|T|=t, i \notin T}[v(T \cup \bar{i}) - v(T)] \\
&\overset{Z=T\cup\bar{i}}{=} \tfrac{((t+1)-1)!(n-(t+1))!}{n!} \sum_{|Z|=t+1, i \in Z}[v(Z) - v(Z \setminus \bar{i})] \\
&\overset{T=Z}{=} \tfrac{((t+1)-1)!(n-(t+1))!}{n!} \sum_{|T|=t+1} \sum_{i \in T}[v(T) - v(T \setminus \bar{i})] = \omega_{t+1}.
\end{aligned}$$

Therefore, $\omega_t \geq \pi_t$ if and only if $\omega_t \geq \omega_{t+1}$, or equivalently, ω_t is a decreasing function of t. In the same fashion, $\omega_t \geq \pi_t$ if and only if $\pi_{t-1} \geq \pi_t$, or equivalently, π_t is a decreasing function of t.

A9. Proof of Theorem 7

If $P_T = \frac{(|T|)!(n-|T|)!}{(n+1)!}$ and v has diminishing marginality, then

$$\begin{aligned}
\sum_{i=1}^{n} \kappa_i[v;\mu] \quad &= \quad \sum_{T \neq \emptyset} \tfrac{(|T|)!(n-|T|)!}{(n+1)!} \sum_{i \in T} [v(T) - v(T \setminus \bar{i})] \\
&\quad - \sum_{Z \neq N} \tfrac{(|Z|)!(n-|Z|)!}{(n+1)!} \sum_{i \notin Z} \big[v(Z \cup \bar{i}) - v(Z)\big] \\
&\overset{T=Z\cup\bar{i}}{=} \quad \sum_{t=1}^{n} \sum_{|T|=t} \tfrac{(|T|)!(n-|T|)!}{(n+1)!} \sum_{i \in T} [v(T) - v(T \setminus \bar{i})] \\
&\quad - \sum_{T \neq \emptyset} \tfrac{(|T|-1)!(n+1-|T|)!}{(n+1)!} \sum_{i \in T} \big[v(T) - v(T \setminus \bar{i})\big] \\
&= \quad \sum_{t=1}^{n} \tfrac{t}{n+1}\omega_t - \sum_{t=1}^{n} \sum_{|T|=t} \tfrac{(|T|-1)!(n+1-|T|)!}{(n+1)!} \sum_{i \in T} \big[v(T) - v(T \setminus \bar{i})\big] \\
&= \quad \sum_{t=1}^{n} \tfrac{t}{n+1}\omega_t - \sum_{t=1}^{n} \tfrac{n+1-t}{n+1}\omega_t = \sum_{t=1}^{n} \tfrac{2t}{n+1}\omega_t - \sum_{t=1}^{n} \omega_t \\
&= \quad n\left[\tfrac{1}{n}\sum_{t=1}^{n}\left(\tfrac{2t}{n+1}\right)(\omega_t) - \left(\tfrac{1}{n}\sum_{t=1}^{n}\omega_t\right)\left(\tfrac{1}{n}\sum_{t=1}^{n}\tfrac{2t}{n+1}\right)\right]
\end{aligned}$$

which is the sample covariance, multiplied by n, between the series $\{\frac{2t}{n+1}\}_{t=1}^{n}$ and the series $\{\omega_t\}_{t=1}^{n}$. As $\frac{2t}{n+1}$ is increasing in t while ω_t is decreasing in t, the sample covariance is non-positive. Therefore, $\sum_{i=1}^{n} \kappa_i[v;\mu] \leq 0$.

If v is super-additive, then $v(\emptyset) = 0$ and v is monotone, i.e., $v(S) \leq v(T)$ if $S \subset T$ (cf Megiddo, 1988). Let

$$\begin{aligned}
\Delta_1 &= \sum_{T \neq \emptyset} (|T|)!(n-|T|)! \sum_{i \in T} \big[v(T) - v(T \setminus \bar{i})\big], \\
\Delta_2 &= \sum_{Z \neq N} (|Z|)!(n-|Z|)! \sum_{i \notin Z} \big[v(Z \cup \bar{i}) - v(Z)\big].
\end{aligned}$$

Note that Δ_1 equals

$$\begin{aligned}
&\sum_{T \neq \emptyset} |T|(|T|)!(n-|T|)!v(T) - \sum_{T \neq \emptyset} \sum_{i \in T} (|T|)!(n-|T|)!v(T \setminus \bar{i}) \\
\overset{Z=T\setminus\bar{i}}{=} &\sum_{T \neq \emptyset} |T|(|T|)!(n-|T|)!v(T) - \sum_{Z \neq N} \sum_{i \notin Z} (|Z|+1)!(n-|Z|-1)!v(Z) \\
= &\sum_{T \neq \emptyset} |T|(|T|)!(n-|T|)!v(T) - \sum_{Z \neq N} (|Z|+1)!(n-|Z|)!v(Z),
\end{aligned}$$

and Δ_2 equals

$$\begin{aligned}
& \sum_{Z\neq N}(|Z|)!(n-|Z|)!\sum_{i\notin Z}v(Z\cup\bar{i}) - \sum_{Z\neq N}(|Z|)!(n-|Z|)!(n-|Z|)v(Z) \\
\overset{T=Z\cup\bar{i}}{=} & \sum_{T\neq\emptyset}(|T|-1)!(n-|T|+1)!\sum_{i\in T}v(T) - \sum_{Z\neq N}(|Z|)!(n-|Z|)!(n-|Z|)v(Z) \\
= & \sum_{T\neq\emptyset}(|T|)!(n-|T|+1)!v(T) - \sum_{Z\neq N}(|Z|)!(n-|Z|)!(n-|Z|)v(Z).
\end{aligned}$$

Thus, $\Delta_1 - \Delta_2$ equals

$$\begin{aligned}
& \sum_{T\neq\emptyset}(|T|)!(n-|T|)!(2|T|-n-1)v(T) + \sum_{Z\neq N}(|Z|)!(n-|Z|)!(n-2|Z|-1)v(Z) \\
= & \; n!(n-1)v(N) - \sum_{T\neq\emptyset,T\neq N}(|T|)!(n-|T|)!v(T) - \sum_{Z\neq N,Z\neq\emptyset}(|Z|)!(n-|Z|)!v(Z) \\
\overset{T=N\setminus Z}{=} & \; n!(n-1)v(N) - \sum_{T\neq\emptyset,T\neq N}(|T|)!(n-|T|)!v(T) - \sum_{T\neq\emptyset,T\neq N}(n-|T|)!(|T|)!v(N\setminus T) \\
= & \; n!(n-1)v(N) - \sum_{T\neq\emptyset,T\neq N}(|T|)!(n-|T|)!\big[v(T)+v(N\setminus T)\big] \\
\geq & \; n!(n-1)v(N) - \sum_{T\neq\emptyset,T\neq N}(|T|)!(n-|T|)!v(N) \\
= & \; n!(n-1)v(N) - \sum_{t=1}^{n-1}\sum_{|T|=t}(|T|)!(n-|T|)!v(N) \\
= & \; n!(n-1)v(N) - \sum_{t=1}^{n-1}n!v(N) = 0.
\end{aligned}$$

Finally, when μ is the benchmark distribution for the Shapley value,

$$\sum_{i=1}^{n}\kappa_i[v;\mu] = \frac{\Delta_1}{(n+1)!} - \frac{\Delta_2}{(n+1)!} = \frac{\Delta_1-\Delta_2}{(n+1)!} \geq 0$$

for the super-additive v.

A10. Proof of Theorem 8

Using $\gamma_i[v;\mu]$ and $\lambda_i[v;\mu]$ in the proof of Theorem 3,

$$\begin{aligned}
\kappa_i[v;\mu] &= \sum_{T\ni i}p^{|T|}(1-p)^{n-|T|}[v(T)-v(T\setminus\bar{i})] \\
&\quad - \sum_{T\ni i}p^{|T|-1}(1-p)^{n-|T|+1}[v(T)-v(T\setminus\bar{i})] \\
&= (2p-1)\sum_{T}p^{|T|-1}(1-p)^{n-|T|}[v(T)-v(T\setminus\bar{i})].
\end{aligned}$$

By Theorem 3, $\alpha = 2p-1$. Finally, by (6),

$$\begin{aligned}
\tilde{\psi}_i[v;\mu] &= (1-(2p-1))\sum_{T\ni i}p^{|T|}(1-p)^{n-|T|}[v(T)-v(T\setminus\bar{i})] \\
&\quad +(1+(2p-1))\sum_{T\ni i}p^{|T|-1}(1-p)^{n-|T|+1}[v(T)-v(T\setminus\bar{i})] \\
&= 4\sum_{T\ni i}p^{|T|}(1-p)^{n-|T|+1}[v(T)-v(T\setminus\bar{i})].
\end{aligned}$$

A11. Proof of Theorem 10

As the ordering τ has a uniform distribution over Ω, each ordering has a probability $\frac{1}{n!}$. Moreover, there are $(|\Xi_i^\tau|)!$ permutations in Ξ_i^τ and $(n-1-|\Xi_i^\tau|)!$ permutations in $N \setminus \Xi_i^\tau \setminus \bar{i}$, the set of elements preceded by i in the ordering τ. Thus, the probability of $\Xi_i^\tau = T$ is $\frac{(|\Xi_i^\tau|)!(n-1-|\Xi_i^\tau|)!}{n!} = \frac{(|T|)!(n-1-|T|)!}{n!}$. Using the double expectation formula, we have

$$
\begin{aligned}
\mathrm{E}[\tilde{\phi}_i^\tau] \quad &= \quad \mathrm{E}\left[\mathrm{E}[\tilde{\phi}_i^\tau \mid \Xi_i^\tau]\right] \\
&= \quad \sum_{T \not\ni i} \mathrm{Prob}(\Xi_i^\tau = T)\mathrm{E}\left[\tilde{\phi}_i^\tau \mid \Xi_i^\tau = T\right] \\
&= \quad \sum_{T \not\ni i} \tfrac{(|T|)!(n-1-|T|)!}{n!} \tfrac{n!(P_T + P_{T\cup\bar{i}})}{(|T|)!(n-|T|-1)!} \left[v(T \cup \bar{i}) - v(T)\right] \\
&= \quad \sum_{T} (P_T + P_{T\cup\bar{i}}) \left[v(T \cup \bar{i}) - v(T)\right] \\
&= \quad \sum_{T} P_T \left[v(T \cup \bar{i}) - v(T)\right] + \sum_{T} P_{T\cup\bar{i}} \left[v(T \cup \bar{i}) - v(T)\right] \\
&\overset{Z=T\cup\bar{i}}{=} \quad \lambda_i[v;\mu] + \sum_{Z} P_Z \left[v(Z) - v(Z \setminus \bar{i})\right] \\
&= \quad \lambda_i[v;\mu] + \gamma_i[v;\mu] = \psi_i[v;\mu].
\end{aligned}
$$

The above proof already implies (9). For the Shapley value, let us apply (4) with $\eta = 0$,

$$
\begin{aligned}
\frac{n!P_{\Xi_i^\tau\cup\bar{i}}}{(|\Xi_i^\tau|)!(n-|\Xi_i^\tau|-1)!} &= \frac{n!\frac{(|\bar{i}\cup\Xi_i^\tau|)!(n-|\bar{i}\cup\Xi_i^\tau|)!}{(n+1)!}}{(|\Xi_i^\tau|)!(n-|\Xi_i^\tau|-1)!} = \frac{1+|\Xi_i^\tau|}{n+1}, \\
\frac{n!P_{\Xi_i^\tau}}{(|\Xi_i^\tau|)!(n-|\Xi_i^\tau|-1)!} &= \frac{n!\frac{(|\Xi_i^\tau|)!(n-|\Xi_i^\tau|)!}{(n+1)!}}{(|\Xi_i^\tau|)!(n-|\Xi_i^\tau|-1)!} = \frac{n-|\Xi_i^\tau|}{n+1}.
\end{aligned}
$$

For the unbiased Shapley value, the right-hand side of (11) is

$$
\begin{aligned}
& 4\mathrm{E}\left\{\mathrm{E}\left[\tfrac{(|\Xi_i^\tau|+1)(n-|\Xi_i^\tau|)}{(n+1)(n+2)}[v(\Xi_i^\tau \cup \bar{i}) - v(\Xi_i^\tau)] \mid \Xi_i^\tau\right]\right\} \\
= \quad & 4\sum_{T \not\ni i} \mathrm{Prob}(\Xi_i^\tau = T)\ \mathrm{E}\left[\tfrac{(|\Xi_i^\tau|+1)(n-|\Xi_i^\tau|)}{(n+1)(n+2)}[v(\Xi_i^\tau \cup \bar{i}) - v(\Xi_i^\tau)] \mid \Xi_i^\tau = T\right] \\
= \quad & 4\sum_{T \not\ni i} \tfrac{(|T|)!(n-1-|T|)!}{n!} \tfrac{(|T|+1)(n-|T|)}{(n+1)n+2)}[v(T \cup \bar{i}) - v(T)] \\
\overset{Z=T\cup\bar{i}}{=} \quad & 4\sum_{Z \ni i} \tfrac{(|Z|)!(n-|Z|+1)!}{(n+2)!}[v(Z) - v(Z \setminus \bar{i})] = \tilde{\Psi}_i[v].
\end{aligned}
$$

Table 2: Multivariate Shapley Value and its Percentage (in Parenthesis)

Variable	R-squared	Squared t-stat	Abs t-stat	MSE: dynamic	MSE: static
$E_{i,t-1}$	0.00331***	390.90***	19.630***	0.0044***	0.0073***
	(42.31)	(48.67)	(29.39)	(17.01)	(27.26)
$E_{i,t-2}$	0.00040*	29.64	5.066*	0.0035**	0.0025*
	(5.15)	(3.69)	(7.58)	(13.57)	(9.45)
$W_{i,t}$	0.00057*	59.76*	7.578**	0.0009	0.0009
	(7.27)	(7.44)	(11.35)	(3.50)	(3.34)
$W_{i,t-1}$	0.00010	12.15	3.088	0.0009	0.0009
	(1.34)	(1.51)	(4.62)	(3.31)	(3.50)
$W_{i,t-2}$	0.00002	1.79	1.056	0.0021*	0.0019*
	(0.20)	(0.22)	(1.58)	(8.30)	(6.95)
$K_{i,t}$	0.00212***	215.47***	14.424***	0.0048***	0.0048***
	(27.15)	(26.82)	(21.60)	(18.69)	(17.78)
$K_{i,t-1}$	0.00077*	55.28*	5.914*	0.0025*	0.0025*
	(9.85)	(6.88)	(8.85)	(9.68)	(9.30)
$K_{i,t-2}$	0.00020	12.93	2.821	0.0026**	0.0022*
	(2.53)	(1.61)	(4.22)	(10.15)	(8.19)
$Y_{i,t}$	0.00015	11.70	3.126	0.0007	0.0007
	(1.86)	(1.46)	(4.68)	(2.57)	(2.58)
$Y_{i,t-1}$	0.00012	9.60	2.568	0.0011	0.0012
	(1.54)	(1.20)	(3.84)	(4.47)	(4.42)
$Y_{i,t-2}$	0.00006	4.03	1.517	0.0023*	0.0019*
	(0.81)	(0.50)	(2.27)	(8.76)	(7.23)

***: 15% or more; **: 10-15%; *: 5-10%

Table 3: Unbiased Multivariate Shapley Value and its Percentage (in Parenthesis)

Variable	R-squared	Squared t-stat	Abs t-stat	MSE: dynamic	MSE: static
$E_{i,t-1}$	0.00319***	387.57***	19.591***	0.0038***	0.0064***
	(43.07)	(48.85)	(29.43)	(19.79)	(30.80)
$E_{i,t-2}$	0.00036	28.70	5.026*	0.0023**	0.0018*
	(4.91)	(3.62)	(7.55)	(12.12)	(8.49)
$W_{i,t}$	0.00057*	60.92*	7.617**	0.0010*	0.0010
	(7.70)	(7.68)	(11.44)	(5.05)	(4.61)
$W_{i,t-1}$	0.00011	12.62	3.121	0.0006	0.0007
	(1.47)	(1.59)	(4.69)	(2.94)	(3.25)
$W_{i,t-2}$	0.00002	1.88	1.072	0.0014*	0.0012*
	(0.22)	(0.24)	(1.61)	(7.26)	(5.92)
$K_{i,t}$	0.00201***	212.08***	14.370***	0.0040***	0.0040***
	(27.07)	(26.73)	(21.59)	(20.78)	(19.20)
$K_{i,t-1}$	0.00068*	52.39*	5.814*	0.0017*	0.0017*
	(9.18)	(6.60)	(8.73)	(8.96)	(8.41)
$K_{i,t-2}$	0.00017	12.26	2.775	0.0018*	0.0015*
	(2.33)	(1.55)	(4.17)	(9.36)	(7.04)
$Y_{i,t}$	0.00013	11.61	3.121	0.0004	0.0005
	(1.82)	(1.46)	(4.69)	(2.29)	(2.38)
$Y_{i,t-1}$	0.00011	9.53	2.565	0.0007	0.0008
	(1.49)	(1.20)	(3.85)	(3.72)	(3.78)
$Y_{i,t-2}$	0.00005	3.82	1.498	0.0015*	0.0013*
	(0.73)	(0.48)	(2.25)	(7.74)	(6.12)
$\hat{\alpha}$	-.227	-.111	-.057	-.508	-.477

***: 15% or more; **: 10-15%; *: 5-10%

Table 4: Trichotomized Marginal Gain, Loss and Effect in the Non-Randomized Data

(U, M, E)	Frequency	$v(U, M, E)$	Marginal Gain	Marginal Loss	Marginal Effect
(-1,-1,-1)	1	5.4173		-11.1971	-11.1971
(-1,-1,0)	2	-5.7798	-11.1971	4.29880	-3.44920
(-1,-1,1)	2	-1.481	4.29880		4.29880
(-1,0,-1)	12	-7.8986		6.97910	6.97910
(-1,0,0)	15	-0.91951	6.97910	-4.31830	1.33040
(-1,0,1)	4	-5.2378	-4.31830		-4.31830
(-1,1,-1)	0	.			
(-1,1,0)	0	.			
(-1,1,1)	2	4.7123			
(0,-1,-1)	1	-4.1197		1.64230	1.64230
(0,-1,0)	5	-2.4774	1.64230	-4.82210	-1.58990
(0,-1,1)	8	-7.2995	-4.82210		-4.82210
(0,0,-1)	45	-2.9165		3.63000	3.63000
(0,0,0)	133	0.7135	3.63000	3.34890	3.48950
(0,0,1)	51	4.0624	3.34890		3.34890
(0,1,-1)	4	-4.6898		2.13620	2.13620
(0,1,0)	8	-2.5536	2.13620	-1.00840	0.56390
(0,1,1)	4	-3.562	-1.00840		-1.00840
(1,-1,-1)	2	2.6511		-0.53780	-0.53780
(1,-1,0)	1	2.1133	-0.53780	-4.87170	-2.70480
(1,-1,1)	1	-2.7584	-4.87170		-4.87170
(1,0,-1)	13	-2.4809		5.72810	5.72810
(1,0,0)	13	3.2472	5.72810	-1.05760	2.33530
(1,0,1)	9	2.1896	-1.05760		-1.05760
(1,1,-1)	1	-8.0136		9.74190	9.74190
(1,1,0)	2	1.7283	9.74190	-0.21330	4.76430
(1,1,1)	1	1.515	-0.21330		-0.21330
Total	340		259	258	338
Weighted			3.04517	2.58613	2.7995