First-Order Efficiency for Probabilistic Value Estimation via A Statistical Viewpoint

Ziqi Liu[1], Kiljae Lee[2], Yuan Zhang[2], and Weijing Tang[1]
[1]*Department of Statistics & Data Science, Carnegie Mellon University*
[2]*Department of Statistics, The Ohio State University*

Abstract

Probabilistic values, including Shapley values and semivalues, provide a model-agnostic framework to attribute the behavior of a black-box model to data points or features, with a wide range of applications including explainable artificial intelligence and data valuation. However, their exact computation requires utility evaluations over exponentially many coalitions, making Monte Carlo approximation essential in modern machine learning applications. Existing estimators are often developed through different identification strategies, including weighted averages, self-normalized weighting, regression adjustment, and weighted least squares. Our key observation is that these seemingly distinct constructions share a common first-order error structure, in which the leading term is an augmented inverse-probability weighted influence term determined by the sampling law and a working surrogate function. This first-order representation yields an explicit expression for the leading mean squared error (MSE), which characterizes how the sampling law and the surrogate jointly determine statistical efficiency. Guided by this criterion, we propose an *Efficiency-Aware Surrogate-adjusted Estimator (EASE)* that directly chooses the sampling law and surrogate to minimize the first-order MSE. We demonstrate that EASE consistently outperforms state-of-the-art estimators for various probabilistic values.

Keywords: Cooperative games, Shapley value, Probabilistic value, Monte Carlo estimation, Surrogate adjustment.

1. Introduction

Shapley values [28], and the broader class of probabilistic values [8, 31], provide an axiomatic framework for quantifying individual players' contributions in a cooperative game. This framework is now widely used in modern machine learning, including model-agnostic data attribution and explainable artificial intelligence [13, 16, 21, 30], where data points or features are treated as players. For a game with player set $[n] = \{1, \ldots, n\}$ and utility function $u : 2^{[n]} \to \mathbb{R}$, where $u(S)$ measures the utility of coalition S, a probabilistic value assigns each player $i \in [n]$ a weighted average of its marginal contributions $u(S \cup \{i\}) - u(S)$ over coalitions $S \subseteq [n] \setminus \{i\}$. For example, $u(S)$ may be the prediction accuracy of a model trained on the data points in S.

Although probabilistic values provide a principled metric for player valuation, their exact computation is intractable: it requires utility evaluations over all 2^n subsets of n, and each evaluation $u(S)$ may be expensive, often involving model retraining or costly inference. The sampling-based approximation has therefore become the standard approach, and a broad literature has developed a range of Monte Carlo estimators for probabilistic values [3, 7]. These methods can be organized by how they identify the target under a sampling distribution. One representative family uses weighted-average identification, which directly represents the target as a weighted average of coalition utilities [5, 9, 14, 18, 30, 32]. Another widely used family exploits weighted least-squares (WLS) identification, which obtains the target through a projection of the utility function onto a structured feature space [4, 10, 19, 21, 24].

Our key observation is that these seemingly different estimator constructions share a common first-order error structure as Monte Carlo sample size $m \to \infty$. For both weighted-average and least-squares estimators, the first-order error term can be written as an augmented inverse-probability-weighted influence term indexed by the sampling law and a working surrogate function. This representation therefore provides a unified first-order characterization of many existing estimators and, more importantly, enables an explicit expression for the leading mean squared error (MSE) in the asymptotic regime $m \to \infty$ with fixed number of players n.

However, for practical probabilistic-value estimation, the primary regime of interest is not an arbitrary $m \to \infty$, but a more restrictive regime $m = O(\text{polynomial}(n))$ due to budget constraint. To address this gap, we establish non-asymptotic remainder bounds for representative weighted-average and WLS estimators. Our results show that the first-order MSE approximation has a small relative error once $m = \Omega(n \operatorname{polylog}(n))$, aligning with the regime studied in the literature [18, 19, 24]. This validation turns the first-order MSE expression into a practical design criterion, which makes explicit how the sampling law and the surrogate choice jointly determine statistical efficiency.

Motivated by this criterion, we propose the *Efficiency-Aware Surrogate-adjusted Estimator (EASE)*, which directly targets the first-order MSE. The optimization problem involves two coupled design choices: fitting the surrogate and choosing the sampling law. Since the objective depends on the unknown utility function, both components must be learned from sampled utility evaluations. Moreover, without structural restrictions, choosing the sampling law would require optimizing over all distributions on $2^{[n]}$, which is computationally infeasible for even moderate n. EASE makes this design problem tractable through a two-stage procedure. The initialization stage uses pilot samples to fit an initial surrogate and construct a residual-aware sampling law within a structured class. The estimation stage then draws samples from this learned law and constructs a cross-fitted augmented inverse-probability-weighted estimator, with the surrogate chosen to minimize the empirical first-order MSE criterion. Numerical results demonstrate that EASE consistently outperforms state-of-the-art estimators for a range of probabilistic-value targets.

2. Probabilistic Values and Existing Estimators

A probabilistic value assigns player i the weighted average

$$\phi_i(u) = \sum_{S \subseteq [n] \setminus \{i\}} \alpha_i^{(n)}(S)\{u(S \cup \{i\}) - u(S)\}, \tag{1}$$

where the weights are nonnegative and sum to one. Semivalues [8] are the special case in which the weights depend only on $|S|$; the Shapley value [28] further sets $\alpha_i^{(n)}(S) = \{n\binom{n-1}{|S|}\}^{-1}$.

Our target is a general linear combination $\tau_{\boldsymbol{a}}(u) = \boldsymbol{a}^\top \boldsymbol{\phi}(u)$, where $\boldsymbol{\phi}(u) = (\phi_1(u), \ldots, \phi_n(u))^\top$ is the probabilistic value vector and $\boldsymbol{a} \in \mathbb{R}^n$ is known. This target class includes individual probabilistic values and other scalar summaries, such as aggregated values for groups of players [17].

In the Monte Carlo setting, we observe sampled coalitions $S_1, \ldots, S_m \sim q$ and their utilities $u(S_t)$, where q is a distribution over subsets of $[n]$. Most existing Monte Carlo estimators can be grouped by how they identify $\tau_{\boldsymbol{a}}(u)$ under this sampling distribution.

Weighted-average identification. Because $\tau_{\boldsymbol{a}}$ is linear in u, there exists a signed coefficient function $\rho_{\boldsymbol{a}}$ such that $\tau_{\boldsymbol{a}}(u) = \sum_{S \subseteq [n]} \rho_{\boldsymbol{a}}(S) u(S)$. Thus, for any sampling distribution q with support containing that of $\rho_{\boldsymbol{a}}$, $\tau_{\boldsymbol{a}}(u)$ can be identified as

$$\tau_{\boldsymbol{a}}(u) = \mathbb{E}_q[\gamma_{\boldsymbol{a},q}(S) u(S)], \qquad \gamma_{\boldsymbol{a},q}(S) := \rho_{\boldsymbol{a}}(S)/q(S). \tag{2}$$

This identity directly yields the Horvitz–Thompson estimator, which replaces the population expectation by a sample mean [5, 9]. With a feasible surrogate h, the same identity also gives a surrogate-adjusted estimator based on a Horvitz–Thompson correction [32]:

$$\hat{\tau}^{\mathrm{HT}}_{\boldsymbol{a},m}(u) = \frac{1}{m}\sum_{t=1}^{m}\gamma_{\boldsymbol{a},q}(S_t)u(S_t), \qquad \hat{\tau}^{\mathrm{sur}}_{\boldsymbol{a},m}(u) = \tau_{\boldsymbol{a}}(h) + \hat{\tau}^{\mathrm{HT}}_{\boldsymbol{a},m}(u-h). \tag{3}$$

Self-normalized and stratified variants [14, 18, 30] further partition the coalition space $\{S : S \subseteq [n]\}$, for example by coalition size $|S|$, estimate the corresponding expectation within each stratum, and aggregate the stratum-level estimates according to their probabilities.

WLS identification. For Shapley values and related semivalues, the target $\tau_{\boldsymbol{a}}(u)$ can also be identified through a population WLS problem:

$$\tau_{\boldsymbol{a}}(u) = \boldsymbol{c}_{\boldsymbol{a}}^{\top}\boldsymbol{\beta}^{\star}(u), \qquad \boldsymbol{\beta}^{\star}(u) \in \arg\min_{\boldsymbol{\beta}} \mathbb{E}_q\left[\frac{w(S)}{q(S)}\{u(S) - \boldsymbol{z}(S)^{\top}\boldsymbol{\beta}\}^2\right], \tag{4}$$

for a feature map $\boldsymbol{z} : 2^{[n]} \to \mathbb{R}^{d_{\boldsymbol{z}}}$, nonnegative weights $w : 2^{[n]} \to \mathbb{R}_+ \cup \{0\}$, and vector $\boldsymbol{c}_{\boldsymbol{a}} \in \mathbb{R}^{d_{\boldsymbol{z}}}$. This identification motivates empirical WLS estimators, which replace the population objective with its sample analogue and plug the resulting coefficient estimate into $\hat{\tau}_{\boldsymbol{a}}(u) = \boldsymbol{c}_{\boldsymbol{a}}^{\top}\hat{\boldsymbol{\beta}}(u)$ [10, 19, 21, 24].

Appendix A collects the explicit forms of these estimator families.

3. A First-Order View of Estimator Design

In this section, we formalize a first-order representation for a broad class of probabilistic-value estimators, including those discussed in Section 2. Specifically, we show that their approximation errors, $\hat{\tau}_{\boldsymbol{a},m}(u) - \tau_{\boldsymbol{a}}(u)$, admit an expansion whose leading term is an augmented weighted average indexed by a surrogate function $h_m : 2^{[n]} \to \mathbb{R}$:

$$\hat{\tau}_{\boldsymbol{a},m}(u) - \tau_{\boldsymbol{a}}(u) = \frac{1}{m}\sum_{t=1}^{m}\psi_{h_m}(S_t; u) + r_m, \qquad \mathbb{E}_q(r_m^2) = o(m^{-1}), \tag{5}$$

where $\psi_h(S; u) = \gamma_{\boldsymbol{a},q}(S)\{u(S) - h(S)\} + \tau_{\boldsymbol{a}}(h) - \tau_{\boldsymbol{a}}(u)$. Within this framework, estimator constructions differ, at the first-order level, in the surrogate h_m that they use or implicitly induce, the class $\mathcal{H}$ to which this surrogate is constrained, and the sampling law q.

For example, the Horvitz–Thompson estimator $\hat{\tau}^{\mathrm{HT}}_{\boldsymbol{a},m}(u)$ corresponds to the degenerate choice $h_m = 0$, with $\mathcal{H} = \{0\}$ and $r_m = 0$. The surrogate-adjusted estimator $\hat{\tau}^{\mathrm{sur}}_{\boldsymbol{a},m}(u)$ also has $r_m = 0$, but uses a specified surrogate h_m chosen from a prescribed class $\mathcal{H}$. Other estimators, such as self-normalized estimators and weighted least-squares estimators, instead induce h_m and $\mathcal{H}$ implicitly. Appendix B and Table 1 give the formal specifications of these explicit or implicit choices for each estimator class.

The first-order expansion (5) immediately gives a structure for mean-squared error (MSE). Consider a vectorized target $\boldsymbol{\tau}_{\boldsymbol{A}}(u) = \boldsymbol{A}^{\top}\boldsymbol{\phi}(u)$, where $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d]$, and a vectorized estimator $\hat{\boldsymbol{\tau}}_{\boldsymbol{A},m}$ collect coordinate estimators for $\tau_{\boldsymbol{a}_1}, \ldots, \tau_{\boldsymbol{a}_d}$. If coordinate j has expansion (5) with surrogate h_j, then the first-order approximation of the MSE is

$$\mathbb{E}_q\|\hat{\boldsymbol{\tau}}_{\boldsymbol{A},m}(u) - \boldsymbol{\tau}_{\boldsymbol{A}}(u)\|_2^2 = \frac{1}{m}\{V(\boldsymbol{A}; q, \boldsymbol{h}) + \mathrm{Rem}_m\}, \qquad \text{with } \mathrm{Rem}_m = o(1),$$

where $\boldsymbol{h} = (h_1, \ldots, h_d)$ and

$$V(\boldsymbol{A}; q, \boldsymbol{h}) = \sum_{j=1}^{d} \mathrm{Var}_q\big[\gamma_{\boldsymbol{a}_j,q}(S)\{u(S) - h_j(S)\}\big]\,. \tag{6}$$

Clearly, $V(\boldsymbol{A}; q, \boldsymbol{h})/m$ is the asymptotic first-order MSE term. It provides a clean understanding how sampling distribution q and explicitly chosen or implicitly induced surrogate $\boldsymbol{h}$ affect the MSE, in an asymptotic regime.

While the asymptotic expansion is informative, practical settings typically restrict m to grow at most polynomially with n, rather than allowing arbitrary $m \to \infty$. We therefore derive finer non-asymptotic guarantees on the sample complexity required to ensure $(1-\epsilon)V(\boldsymbol{A}; q, \boldsymbol{h}) \leq \mathrm{MSE} \leq (1+\epsilon)V(\boldsymbol{A}; q, \boldsymbol{h})$. For representative self-normalized estimators, including OFA [18] and Stratified SVARM [14], as well as WLS estimators such as KernelSHAP [4, 21] and LeverageSHAP [24], we show that a sample size of $O(m\,\mathrm{polylog}(n)/\epsilon^2)$ suffices for $V(\boldsymbol{A}; q, \boldsymbol{h})$ to dominate the remainder term in the MSE. This matches the primary regime of interest in the literature [18, 19, 24]. Details are provided in Appendix C.

This theoretical understanding suggests the optimal choice of q and $\boldsymbol{h}$ is to minimize $V(\boldsymbol{A}; q, \boldsymbol{h})$ when designing practical estimation algorithms. However, existing estimators often construct surrogate adjustments using criteria intrinsic to their procedures, such as self-normalization, weighted least squares. Such choices do not necessarily minimize the first-order MSE criterion.

4. EASE: Efficiency-Aware Surrogate Estimation

EASE directly targets the first-order MSE criterion (6). Our basic construction is the following augmented inverse-probability weighted (AIPW) estimator:

$$\hat{\tau}^{\mathrm{AIPW}}_{\boldsymbol{a}_j}(u) = \tau_{\boldsymbol{a}_j}(h_j) + \frac{1}{m}\sum_{t=1}^{m} \gamma_{\boldsymbol{a}_j,q}(S_t)\{u(S_t) - h_j(S_t)\}, \quad j = 1, \ldots, d, \tag{7}$$

where h_j is obtained via cross-fitting. The AIPW construction is closely related to the regression-adjusted estimator of Witter et al. [32]. A key advantage is that its MSE is exactly $V(\boldsymbol{A}; q, \boldsymbol{h})/m$, matching the leading-order MSE of any regular estimator with the same $(q, \boldsymbol{h})$. Consequently, the AIPW estimator with $(q, \boldsymbol{h})$ optimized for the criterion $V(\boldsymbol{A}; q, \boldsymbol{h})$ achieves first-order efficiency among all estimators satisfying the expansion (5). This optimization problem involves two coupled design choices: how to construct the surrogate h, and how to choose the sampling law q.

Surrogate fitting. To minimize $V(\boldsymbol{A}; q, \boldsymbol{h})$ under a fixed q, we note that the quantity in (6) can be written as a profiled squared-loss objective. Introducing an auxiliary centering vector $\boldsymbol{\mu}$, we define

$$\mathscr{V}(\boldsymbol{A}; q, \boldsymbol{h}, \boldsymbol{\mu}) = \sum_{j=1}^{d} \mathbb{E}_q\big[\{\gamma_{\boldsymbol{a}_j,q}(S)(u(S) - h_j(S)) - \mu_j\}^2\big]\,. \tag{8}$$

Then, $V(\boldsymbol{A}; q, \boldsymbol{h}) = \min_{\boldsymbol{\mu}\in\mathbb{R}^d} \mathscr{V}(\boldsymbol{A}; q, \boldsymbol{h}, \boldsymbol{\mu})$, where the minimizer centers each term at its mean. Therefore, given the observed samples indexed by $\mathcal{T} \subseteq [m]$, we can estimate the optimal $\boldsymbol{h}$ by the empirical least-squares problem jointly on $(\boldsymbol{h}, \boldsymbol{\mu})$:

$$(\hat{\boldsymbol{h}}_{\mathcal{T}}, \hat{\boldsymbol{\mu}}_{\mathcal{T}}) \in \operatorname*{arg\,min}_{\boldsymbol{h}\in\mathcal{H},\ \boldsymbol{\mu}\in\mathbb{R}^d} \sum_{t\in\mathcal{T}}\sum_{j=1}^{d} \big[\gamma_{\boldsymbol{a}_j,q}(S_t)\{Y_t - h_j(S_t)\} - \mu_j\big]^2\,. \tag{9}$$

The fitted surrogate $\hat{\boldsymbol{h}}_{\mathcal{T}}$ is then used in the AIPW estimator. The objective for $\boldsymbol{h}$ is thus directly tailored to minimize the first-order MSE, rather than merely to approximate u uniformly as in [32].

Residual-aware sampling. The first-order MSE criterion also provides guidance for choosing q. Since optimizing over all distributions on $2^{[n]}$ is infeasible even for moderate n, we restrict q to a structured class. Specifically, we consider sampling laws that are constant on a pre-specified partition $\mathscr{C} = \{C_1, \ldots, C_K\}$, such as the partition by coalition size. For a fixed surrogate $\boldsymbol{h} = (h_1, \ldots, h_d)$, define the cellwise average squared residual

$$M_k(\boldsymbol{h}) := \frac{1}{|C_k|} \sum_{S \in C_k} \sum_{j=1}^{d} \rho_{\boldsymbol{a}_j}(S)^2 \{u(S) - h_j(S)\}^2, \quad k = 1, \ldots, K. \tag{10}$$

The oracle sampling law is then given by $q^\star(S; \boldsymbol{h}) = \sqrt{M_k(\boldsymbol{h})} / \sum_{r=1}^{K} |C_r| \sqrt{M_r(\boldsymbol{h})}$ for $S \in C_k$. This corresponds to a Neyman-type allocation [26]: cells with larger average squared residuals $M_k(\boldsymbol{h})$ after surrogate adjustment receive more samples.

Because $M_k(\boldsymbol{h})$ is unknown before sampling, EASE adopts a two-stage plug-in procedure. First, EASE draws pilot coalitions from an initialization law $q^{\text{init}}(S) \propto \sqrt{M_k^{\text{init}}}$ for $S \in C_k$, where $M_k^{\text{init}} := \frac{1}{|C_k|} \sum_{S \in C_k} \sum_{j=1}^{d} \rho_{\boldsymbol{a}_j}(S)^2$. EASE then fits an initial surrogate using the empirical version of (8), estimates the cellwise risks $M_k(\hat{\boldsymbol{h}})$, and sets $q^{\text{final}}(S) \propto \sqrt{\hat{M}_k}$ for $S \in C_k$. It then draws final samples from q^{final}, refits the surrogate on the training folds, and evaluates AIPW (7) on the held-out folds. Appendix D provides full pseudocode and implementation details.

5. Experiments

We evaluate EASE on synthetic structured sum-of-unanimity (SOU) games, a standard testbed for Shapley-value and probabilistic-value estimators [17, 18]. Exact probabilistic values for SOU games are available in closed form, allowing accurate evaluation of estimation error. All main-text results use the same fixed SOU benchmark with $n = 40$ players, where the goal is to estimate the vector of individual values $\boldsymbol{\phi}$. Appendix F provides detailed game construction, additional SOU settings, and real-world feature-attribution benchmarks for ViT on ImageNet [6], CIFAR-10 [15], BreastCancer [29], NHANES I [25], and Communities & Crime [27]. Code is available at https://github.com/zqliu01/ease_probabilistic_values.

Matched comparisons isolate the design effect. We compare EASE with three representative baselines while matching the working surrogate class used explicitly or implicitly by each baseline. For RegressionMSR [32], we use a linear player-indicator class. For OFA [18], we match its player-by-size implicit class. For PolySHAP [10], we match the degree-two polynomial class. The details of the surrogate classes are in Appendix F. Figure 1 summarizes the relative estimation errors $\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\|_2^2 / \|\boldsymbol{\phi}\|_2^2$ against the average number of utility evaluations per player. EASE achieves lower relative squared error at every measured budget in all three matched comparisons. These results demonstrate the benefits of our efficiency-aware design: using the same working surrogate class, the efficiency-guided selection of $\boldsymbol{h}$ and q yields consistent improvements over existing estimators.

Benchmark Across Various Probabilistic Values. Figure 2 benchmarks EASE on Shapley, two Beta-Shapley targets $(4, 1)$ and $(1, 4)$ [16], and three weighted Banzhaf targets with $p = 0.25$, 0.5, and 0.75 [30]. We compare EASE-FO, which uses player indicators, and EASE-SP, which uses player-by-size indicators, against the baseline suite in Appendix F. To quantify estimator performance, we use *area under the convergence curve* (AUCC), following recent works [17, 18]. AUCC

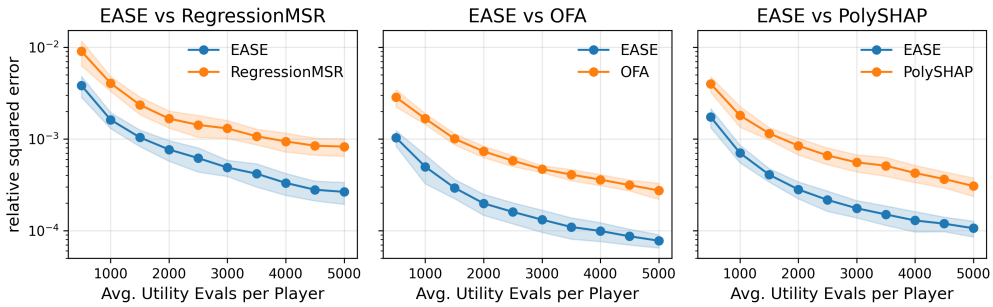


Figure 1: Matched Shapley comparisons on the fixed SOU benchmark ($n = 40$). Each panel matches the working surrogate class; lower log-scale relative squared error is better. Curves show mean ± 1 s.d. over 10 runs.

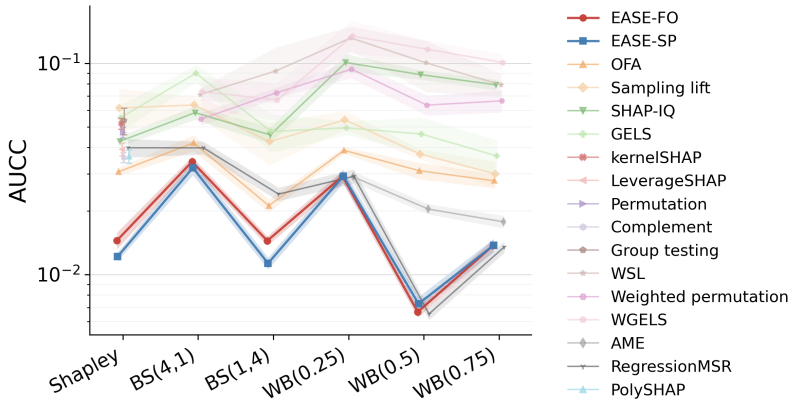


Figure 2: AUCC benchmark on the fixed SOU benchmark ($n = 40$). Lower is better; log scale; mean ± 1 s.d. over 10 runs.

averages relative squared error over budgets $\{50, 100, \ldots, 5000\}$ and lower AUCC values indicate better sample efficiency. Figure 2 shows that, for Beta-Shapley values (including the Shapley values), both EASE variants outperform all baselines, often by a significant margin, with EASE-SP typically providing the best performance. For weighted Banzhaf values, while AME and RegressionMSR also perform well, EASE remains highly competitive.

Overall, these results demonstrate that, by learning the surrogate $\boldsymbol{h}$ and choosing the sampling law q to minimize first-order MSE, the EASE framework offers consistently strong performance across a diverse range of probabilistic values.

6. Conclusion

In this paper, we study the common first-order error structure underlying a broad class of Monte Carlo estimators for probabilistic values. Motivated by this perspective, we propose EASE, whose central principle is that the choice of both the surrogate and the sampling law should be efficiency-guided. We demonstrate both theoretically and empirically that EASE consistently improves upon a wide range of existing estimators.

References

[1] Javier Castro, Daniel Gómez, and Juan Tejada. Polynomial calculation of the shapley value based on sampling. *Computers & Operations Research*, 36(5):1726–1730, 1 May 2009.

[2] Javier Castro, Daniel Gómez, Elisenda Molina, and Juan Tejada. Improving polynomial estimation of the shapley value by stratified random sampling with optimum allocation. *Computers & Operations Research*, 82:180–188, 1 June 2017.

[3] Hugh Chen, Ian C Covert, Scott M Lundberg, and Su-In Lee. Algorithms to estimate shapley value feature attributions. *Nature Machine Intelligence*, 5(6):590–601, 1 June 2023.

[4] Tyler Chen, Akshay Seshadri, Mattia Jacopo Villani, Pradeep Niroula, Shouvanik Chakrabarti, Archan Ray, Pranav Deshpande, Romina Yalovetzky, Marco Pistoia, and Niraj Kumar. A unified framework for provably efficient algorithms to estimate shapley values. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[5] Ian Covert and Su-In Lee. Improving KernelSHAP: Practical shapley value estimation using linear regression. In Arindam Banerjee and Kenji Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 3457–3465. PMLR, 2021. URL https://proceedings.mlr.press/v130/covert21a.html.

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[7] Junwei Deng, Yuzheng Hu, Pingbang Hu, Ting-Wei Li, Shixuan Liu, Jiachen T Wang, Dan Ley, Qirun Dai, Benhao Huang, Jin Huang, Cathy Jiao, Hoang Anh Just, Yijun Pan, Jingyan Shen, Yiwen Tu, Weiyi Wang, Xinhe Wang, Shichang Zhang, Shiyuan Zhang, Ruoxi Jia, Himabindu Lakkaraju, Hao Peng, Weijing Tang, Chenyan Xiong, Jieyu Zhao, Hanghang Tong, Han Zhao, and Jiaqi W Ma. A survey of data attribution: Methods, applications, and evaluation in the era of generative AI. 29 August 2025.

[8] Pradeep Dubey, Abraham Neyman, and Robert James Weber. Value theory without efficiency. *Math. Oper. Res.*, 6(1):122–128, February 1981. doi: 10.1287/moor.6.1.122. URL http://dx.doi.org/10.1287/moor.6.1.122.

[9] Fabian Fumagalli, Maximilian Muschalik, Patrick Kolpaczki, Eyke Hüllermeier, and Barbara Hammer. SHAP-IQ: Unified approximation of any-order shapley interactions. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[10] Fabian Fumagalli, R Teal Witter, and Christopher Musco. PolySHAP: Extending KernelSHAP with interaction-informed polynomial regression. In *International Conference on Learning Representations*, 2026.

[11] Jaroslav Hájek. Comment on "an essay on the logical foundations of survey sampling, part one". In V. P. Godambe and D. A. Sprott, editors, *Foundations of Statistical Inference*, page 236. Holt, Rinehart and Winston, Toronto, 1971.

[12] D G Horvitz and D J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1 December 1952.

[13] R Jia, David Dao, Boxin Wang, F Hubis, Nicholas Hynes, Nezihe Merve Gürel, Bo Li, Ce Zhang, D Song, and C Spanos. Towards efficient data valuation based on the shapley value. *AISTATS*, abs/1902.10275:1167–1176, 27 February 2019. URL https://proceedings.mlr.press/v89/jia19a.html.

[14] Patrick Kolpaczki, Viktor Bengs, Maximilian Muschalik, and Eyke Hüllermeier. Approximating the shapley value without marginal contributions. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(12):13246–13255, 24 March 2024. doi: 10.1609/aaai.v38i12.29225. URL http://dx.doi.org/10.1609/aaai.v38i12.29225.

[15] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

[16] Yongchan Kwon and James Y Zou. Beta shapley: A unified and noise-reduced data valuation framework for machine learning. *International Conference on Artificial Intelligence and Statistics*, 151:8780–8802, 26 October 2021. URL https://proceedings.mlr.press/v151/kwon22a.

[17] Kiljae Lee, Ziqi Liu, Weijing Tang, and Yuan Zhang. Faithful group shapley value. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=z6d5MRMDNf.

[18] Weida Li and Yaoliang Yu. One sample fits all: Approximating all probabilistic values simultaneously and efficiently. *Neural Information Processing Systems*, abs/2410.23808, 31 October 2024.

[19] Weida Li and Yaoliang Yu. Faster approximation of probabilistic and distributional values via least squares. In B Kim, Y Yue, S Chaudhuri, K Fragkiadaki, M Khan, and Y Sun, editors, *International Conference on Representation Learning*, volume 2024, pages 51182–51216, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/df22a19686a558e74f038e6277a51f68-Paper-Conference.pdf.

[20] Jinkun Lin, Anqi Zhang, Mathias Lécuyer, Jinyang Li, Aurojit Panda, and Siddhartha Sen. Measuring the effect of training data on deep learning predictions via randomized experiments. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 13468–13504. PMLR, 2022.

[21] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Neural Information Processing Systems*, pages 4765–4774, 22 May 2017. doi: 10.5555/3295222.3295230. URL http://dx.doi.org/10.5555/3295222.3295230.

[22] Sasan Maleki, Long Tran-Thanh, Greg Hines, Talal Rahwan, and Alex Rogers. Bounding the estimation error of sampling-based shapley value approximation. *arXiv [cs.GT]*, 18 June 2013.

[23] Nicholas Moehle, Stephen Boyd, and Andrew Ang. Portfolio performance attribution via shapley value, 2021. URL http://arxiv.org/abs/2102.05799.

[24] Christopher Musco and R T Witter. Provably accurate shapley value estimation via leverage score sampling. *ArXiv*, abs/2410.01917, 2 October 2024. doi: 10.48550/arXiv.2410.01917. URL https://ui.adsabs.harvard.edu/abs/2024arXiv241001917M/abstract.

[25] National Center for Health Statistics. Plan and operation of the health and nutrition examination survey, united states, 1971–1973. Technical Report Vital and Health Statistics, Series 1, No. 10a, U.S. Department of Health, Education, and Welfare, 1975.

[26] Jerzy Neyman. On the two different aspects of the representative method : The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97(4):558–606, 1934.

[27] Michael Redmond and Alok Baveja. A data-driven software tool for enabling cooperative information sharing among police departments. *European Journal of Operational Research*, 141(3):660–678, 2002.

[28] Lloyd S Shapley. A value for n-person games. In Harold W Kuhn and Albert W Tucker, editors, *Contributions to the Theory of Games II*, number 28 in Annals of Mathematics Studies, pages 307–317. Princeton University Press, 1953.

[29] W. Nick Street, William H. Wolberg, and Olvi L. Mangasarian. Nuclear feature extraction for breast tumor diagnosis. In *IS&T/SPIE International Symposium on Electronic Imaging: Science and Technology*, pages 861–870, 1993.

[30] Jiachen T Wang and R Jia. Data banzhaf: A robust data valuation framework for machine learning. *AISTATS*, 206:6388–6421, 30 May 2022. URL https://proceedings.mlr.press/v206/wang23e.

[31] Robert James Weber. Probabilistic values for games. In Alvin E Roth, editor, *The Shapley Value: Essays in Honor of Lloyd S. Shapley*, pages 101–120. Cambridge University Press, Cambridge, 1988.

[32] R Teal Witter, Yurong Liu, and Christopher Musco. Regression-adjusted monte carlo estimators for shapley values and probabilistic values. In *Neural Information Processing Systems*, 2025.

[33] Mengmeng Wu, Ruoxi Jia, Changle Lin, Wei Huang, and Xiangyu Chang. Variance reduced shapley value estimation for trustworthy data valuation. *Computers & Operations Research*, 159:106305, 2023. doi: 10.1016/j.cor.2023.106305. URL https://www.sciencedirect.com/science/article/pii/S0305054823001697.

[34] Jiayao Zhang, Qiheng Sun, Jinfei Liu, Li Xiong, Jian Pei, and Kui Ren. Efficient sampling approaches to shapley value approximation. *Proceedings of the ACM on Management of Data*, 1(1):1–24, 26 May 2023. doi: 10.1145/3588728. URL http://dx.doi.org/10.1145/3588728.

Camera-Ready Edits Summary

We thank the reviewers for their constructive comments. In the camera-ready version, we made several revisions to improve the empirical scope and clarify the behavior of EASE across different problem sizes and surrogate classes. The main substantive changes are in Appendix Section F.6, *Additional Benchmark Results for SOU Games*, and Appendix Section F.7, *Benchmark Results for Real-World Feature-Attribution Tasks*. We also added a short pointer in Section 5 in the main text to direct readers to these additional experiments.

- To address the reviewers' comments on broader empirical validation beyond the synthetic SOU benchmark, we added Appendix Section F.7, which evaluates EASE on five real-world feature-attribution tasks: ViT on ImageNet, CIFAR-10, BreastCancer, NHANES I, and Communities & Crime. This section reports the dataset sources, number of players/features, prediction models, and the AUCC evaluation protocol with 10 budget checkpoints up to $m = 40n$ utility evaluations.
- To assess scalability across game sizes and structural regimes, we expanded the SOU benchmark results in Appendix Section F.6 to include player counts $n \in \{40, 80, 160\}$ and structural parameters $\eta \in \{0.25, 0.5, 0.75\}$.
- We revised the appendix discussion to clarify the empirical behavior of the two EASE variants. In particular, the new discussion highlights that EASE-FO remains among the strongest methods across both SOU and real-world settings, while EASE-SP can be more variable because its richer player-by-size surrogate has order-n^2 dimensions and is therefore harder to estimate accurately when the utility evaluation budget scales only linearly with n.

Appendix

The appendices are organized to support the compressed main text. Appendix A gives explicit estimator forms and notation. Appendix B states the common first-order representation. Appendix C explains when the first-order MSE approximation is accurate at polynomial budgets. Appendix D gives the EASE procedure. Appendix E records an extension to bundled sampling, and Appendix F gives experimental construction, additional SOU results, and real-world benchmark results.

Appendix A. Estimator Families and Notation

This appendix spells out the estimator families named in the main text. The goal is to make the short terminology in Sections 2–3 traceable to explicit estimators.

A.1. Coefficient Form of Probabilistic Values

For player i, define the signed coefficient

$$\rho_i(S) := \alpha_i^{(n)}(S \setminus \{i\})\mathbf{1}\{i \in S\} - \alpha_i^{(n)}(S)\mathbf{1}\{i \notin S\}, \qquad S \subseteq [n], \tag{11}$$

where the first term is interpreted as zero if $i \notin S$ and the second as zero if $i \in S$. Then

$$\phi_i(u) = \sum_{S \subseteq [n]} \rho_i(S)u(S), \qquad \rho_{\boldsymbol{a}}(S) := \sum_{i=1}^{n} a_i \rho_i(S), \qquad \tau_{\boldsymbol{a}}(u) = \sum_{S \subseteq [n]} \rho_{\boldsymbol{a}}(S)u(S).$$

Under a sampling law q, the inverse-weight coefficient is

$$\gamma_{\boldsymbol{a},q}(S) = \rho_{\boldsymbol{a}}(S)/q(S),$$

provided $q(S) > 0$ whenever $\rho_{\boldsymbol{a}}(S) \neq 0$.

A.2. Weighted-Average and Horvitz–Thompson Estimators

The direct inverse-weighted estimator is the Horvitz–Thompson estimator [12]:

$$\hat{\tau}^{\mathrm{HT}}_{\boldsymbol{a},m}(u) = \frac{1}{m} \sum_{t=1}^{m} \gamma_{\boldsymbol{a},q}(S_t)u(S_t). \tag{12}$$

In the probabilistic-value literature, this includes weighted sample-mean estimators such as the unbiased weighted KernelSHAP variant [5] and order-one SHAP-IQ [9]. The estimator is unbiased but can have high variance when $|\rho_{\boldsymbol{a}}(S)|/q(S)$ is large.

A.3. AIPW and Regression-Adjusted Estimators

For a computable surrogate h, the AIPW estimator is

$$\hat{\tau}^{\mathrm{AIPW}}_{\boldsymbol{a},m}(u) = \tau_{\boldsymbol{a}}(h) + \frac{1}{m} \sum_{t=1}^{m} \gamma_{\boldsymbol{a},q}(S_t)\{u(S_t) - h(S_t)\}. \tag{13}$$

This is the regression-adjusted form used by RegressionMSR [32]. The surrogate is not assumed to be correct; it is a variance-reduction device. If h tracks u well on coalitions with large weights, the residual correction in (13) has lower variance than (12).

Table 1: How the main estimator families enter the first-order view.

Family	Representative methods	First-order surrogate class
Weighted average	HT [12], unbiased weighted KernelSHAP variant [5], SHAP-IQ [9]	$\{0\}$
Regression adjustment	RegressionMSR [32], AIPW variants	explicit $\mathcal{H}^{\mathrm{sur}}$
Self-normalization	Banzhaf [30], Stratified SVARM [14], OFA [18]	partition-induced cellwise class
WLS	KernelSHAP [21], LeverageSHAP [24], GELS/WGELS [19], PolySHAP [10]	feature span $\{\boldsymbol{z}^\top \boldsymbol{\beta}\}$

A.4. Self-Normalized and Post-Stratified Estimators

Let $\mathscr{C} = \{C_1, \ldots, C_K\}$ be a partition of $2^{[n]}$, let $\pi_k = \mathbb{P}_q(S \in C_k)$, and let $N_k = \sum_{t=1}^m \mathbf{1}\{S_t \in C_k\}$. Rewriting (12) by cells gives

$$\hat{\tau}^{\mathrm{HT}}_{\boldsymbol{a},m}(u) = \sum_{k:N_k>0} \frac{N_k}{m} \left\{ \frac{1}{N_k} \sum_{t:S_t \in C_k} \gamma_{\boldsymbol{a},q}(S_t) u(S_t) \right\}.$$

The self-normalized or Hájek-type estimator [11] replaces the random cell frequency N_k/m with the known design probability π_k:

$$\hat{\tau}^{\mathrm{Hajek}}_{\boldsymbol{a},m}(u) = \sum_{k:N_k>0} \pi_k \left\{ \frac{1}{N_k} \sum_{t:S_t \in C_k} \gamma_{\boldsymbol{a},q}(S_t) u(S_t) \right\}. \tag{14}$$

The Banzhaf estimator of Wang and Jia [30] uses player-in/player-out cells. Stratified SVARM [14] further splits by coalition size. OFA [18] uses size-indexed normalization for broader probabilistic values.

A.5. Weighted Least-Squares Estimators

WLS estimators identify the target through a population projection and estimate it by empirical regression. In the notation below, $\boldsymbol{z}(S)$ is the feature map, $w(S)$ is the WLS weight, and $\boldsymbol{c}_{\boldsymbol{a}}$ is the readout vector for target $\tau_{\boldsymbol{a}}$. A ridge-stabilized empirical version is

$$\hat{\tau}^{\mathrm{WLS},\lambda}_{\boldsymbol{a},m}(u) = \boldsymbol{c}_{\boldsymbol{a}}^\top \hat{\boldsymbol{\beta}}_{m,\lambda}, \qquad \hat{\boldsymbol{\beta}}_{m,\lambda} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{t=1}^m \frac{w(S_t)}{q(S_t)} \{u(S_t) - \boldsymbol{z}(S_t)^\top \boldsymbol{\beta}\}^2 + \lambda \|\boldsymbol{\beta}\|_2^2 \right\}. \tag{15}$$

KernelSHAP [21], LeverageSHAP [24], GELS/WGELS [19], and PolySHAP [10] differ in w, $\boldsymbol{z}$, $\boldsymbol{c}_{\boldsymbol{a}}$, and q, but they all fit the template in (15).

Some method names have multiple variants in the literature. The table classifies the estimator form used in the first-order map, not the full paper or software package associated with a method name.

Appendix B. First-Order Unification

This appendix gives the formal version of the unification summarized in Section 3. Throughout, $S_1, \ldots, S_m \sim q$ are i.i.d.

Definition 1 (Regular estimator sequence) *An estimator sequence $\{\hat{\tau}_{\boldsymbol{a},m}\}_{m\geq 1}$ is* regular *relative to a class $\mathcal{H}$ if there exist functions $h_m \in \mathcal{H}$ such that, conditional on any training data used to choose h_m, h_m is fixed and independent of the evaluation samples, and*

$$\hat{\tau}_{\boldsymbol{a},m}(u) - \tau_{\boldsymbol{a}}(u) = \frac{1}{m}\sum_{t=1}^{m} \psi_{h_m}(S_t; u) + r_m, \qquad \mathbb{E}_q(r_m^2) = o(m^{-1}),$$

where ψ_h is defined as

$$\psi_h(S; u) = \gamma_{\boldsymbol{a},q}(S)\{u(S) - h(S)\} + \tau_{\boldsymbol{a}}(h) - \tau_{\boldsymbol{a}}(u)$$

HT and AIPW. The HT estimator in (12) is regular relative to $\mathcal{H} = \{0\}$ with $r_m = 0$. The AIPW estimator in (13) is regular relative to any explicit class $\mathcal{H}^{\text{sur}}$ on which $h \mapsto \tau_{\boldsymbol{a}}(h)$ is computable. For a fixed surrogate the statement is unconditional; for a cross-fitted surrogate it holds fold by fold after conditioning on the training folds, which is the independence used in Section 4.

Self-normalized estimators induce cellwise surrogates. For the partition $\mathscr{C}$, define

$$\mathcal{H}_{\mathscr{C},\gamma_{\boldsymbol{a},q}} := \left\{ h : \gamma_{\boldsymbol{a},q}(S)h(S) = \sum_{k=1}^{K} \omega_k \mathbf{1}\{S \in C_k\},\ \omega_k \in \mathbb{R} \right\}.$$

If $\gamma_{\boldsymbol{a},q}$ is either identically zero or everywhere nonzero on each cell, then the Hájek estimator in (14) is regular relative to $\mathcal{H}_{\mathscr{C},\gamma_{\boldsymbol{a},q}}$. The induced surrogate is the cellwise conditional mean of the weighted utility:

$$h_{\mathscr{C}}(S) = \frac{\mathbb{E}_q[\gamma_{\boldsymbol{a},q}(S)u(S) \mid S \in C_k]}{\gamma_{\boldsymbol{a},q}(S)}, \qquad S \in C_k,$$

on cells where $\gamma_{\boldsymbol{a},q}$ is nonzero. Thus self-normalization is a form of implicit residual adjustment.

WLS estimators induce projection surrogates. Let

$$\mathcal{H}_{\text{WLS}} = \{h : h(S) = \boldsymbol{z}(S)^\top \boldsymbol{\beta},\ \boldsymbol{\beta} \in \mathbb{R}^{d_{\boldsymbol{z}}}\}.$$

Let h_w be the population WLS projection of u onto this feature span under weight w. Under the support, leverage, spectrum, and ridge-stability conditions used in Appendix C, the ridge WLS estimator in (15) admits

$$\hat{\tau}^{\text{WLS},\lambda}_{\boldsymbol{a},m}(u) - \tau_{\boldsymbol{a}}(u) = \frac{1}{m}\sum_{t=1}^{m} \psi_{h_w}(S_t; u) + r_m, \qquad \mathbb{E}_q(r_m^2) = o(m^{-1}).$$

The WLS feature span is therefore the implicit surrogate class.

Explicit and implicit surrogates compose. If an estimator $\hat{\tau}_{\boldsymbol{a},m}$ is regular relative to an implicit class $\mathcal{H}^{\text{imp}}$, then the residual estimator

$$\hat{\tau}^{\text{sur}}_{\boldsymbol{a},m}(u) := \tau_{\boldsymbol{a}}(h^{\text{sur}}) + \hat{\tau}_{\boldsymbol{a},m}(u - h^{\text{sur}})$$

is regular relative to the sum class

$$\mathcal{H}^{\text{sur}} + \mathcal{H}^{\text{imp}} = \{h^{\text{sur}} + h^{\text{imp}} : h^{\text{sur}} \in \mathcal{H}^{\text{sur}}, h^{\text{imp}} \in \mathcal{H}^{\text{imp}}\}.$$

This explains why explicit regression adjustment can be layered on top of self-normalized or WLS constructions.

Appendix C. First-Order MSE and Remainder Support

This appendix supports the polynomial-budget statement in Section 3. Let $\boldsymbol{\tau}_{\boldsymbol{A}}(u) = \boldsymbol{A}^\top \boldsymbol{\phi}(u)$, where $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d]$. The purpose is not to prove a uniform finite-sample guarantee for every adaptive estimator. It is to show, for representative self-normalized and WLS estimators, that the first-order variance used for design dominates the remainder once the sampling budget is polynomial in n.

C.1. Vector MSE Expansion

Suppose each coordinate has the regular expansion

$$\hat{\tau}_{\boldsymbol{a}_j,m}(u) - \tau_{\boldsymbol{a}_j}(u) = \frac{1}{m}\sum_{t=1}^{m} \psi_{h_j}(S_t; u) + r_{m,j}.$$

Let $\boldsymbol{r}_m = (r_{m,1}, \ldots, r_{m,d})$. If

$$V(\boldsymbol{A}; q, \boldsymbol{h}) = \sum_{j=1}^{d} \mathrm{Var}_q[\gamma_{\boldsymbol{a}_j,q}(S)\{u(S) - h_j(S)\}]$$

is nonzero, then

$$\left| \frac{\mathbb{E}_q \|\hat{\boldsymbol{\tau}}_{\boldsymbol{A},m}(u) - \boldsymbol{\tau}_{\boldsymbol{A}}(u)\|_2^2}{V(\boldsymbol{A}; q, \boldsymbol{h})/m} - 1 \right| \le 2\sqrt{\frac{m\mathbb{E}_q\|\boldsymbol{r}_m\|_2^2}{V(\boldsymbol{A}; q, \boldsymbol{h})}} + \frac{m\mathbb{E}_q\|\boldsymbol{r}_m\|_2^2}{V(\boldsymbol{A}; q, \boldsymbol{h})}. \tag{16}$$

Thus the first-order term $V(\boldsymbol{A}; q, \boldsymbol{h})/m$ is a useful design objective when the scaled remainder is small relative to $V(\boldsymbol{A}; q, \boldsymbol{h})$.

C.2. Self-Normalized Remainder Scale

For the Hájek estimator in (14), let K_+ be the number of cells with positive probability, let $\pi_{\min}$ be the smallest positive cell probability, let

$$\Gamma_{2,\mathscr{C}}^2 = \frac{1}{K_+} \sum_{k:\pi_k>0} \mathbb{E}_q[\gamma_{\boldsymbol{a},q}(S)^2 \mid S \in C_k], \qquad \Gamma_\infty = \sup_{S:q(S)>0} |\gamma_{\boldsymbol{a},q}(S)|.$$

Under the cell compatibility condition in Appendix B,

$$\mathbb{E}_q(r_m^2) \le C_1 \|u\|_\infty^2 \left\{ \frac{K_+ \Gamma_{2,\mathscr{C}}^2}{m^2} + \Gamma_\infty^2 \exp(-C_2 m \pi_{\min}) \right\}. \tag{17}$$

For a fixed Shapley-value coordinate, the membership-size partition used by OFA and Stratified SVARM has $K_+ = O(n)$. Substituting their size allocations into (17) gives

$$\mathbb{E}_q(r_m^2) \le C_1 \|u\|_\infty^2 \begin{cases} \dfrac{n \log n}{m^2} + n \exp(-C_2 m / n^{3/2}), & \text{OFA}, \\ \dfrac{n \log^2 n}{m^2} + \log^2 n \, \exp(-C_2 m/(n \log n)), & \text{Stratified SVARM}. \end{cases} \tag{18}$$

These bounds yield the $m = \Omega(n \log n/\epsilon^2)$ and $m = \Omega(n \log^2 n/\epsilon^2)$ scales quoted in the main text for relative tolerance ϵ when the first-order variance is nondegenerate.

C.3. WLS Remainder Scale

For WLS estimators, define the weighted Gram matrix and leverage score

$$\boldsymbol{G} = \mathbb{E}_q[\tilde{w}(S)\boldsymbol{z}(S)\boldsymbol{z}(S)^\top], \qquad \ell(S) = \tilde{w}(S)\boldsymbol{z}(S)^\top \boldsymbol{G}^+ \boldsymbol{z}(S), \qquad \tilde{w}(S) = w(S)/q(S),$$

and let $L = \sup_S \ell(S)$. Here $\boldsymbol{G}^+$ is the Moore–Penrose pseudoinverse and the support condition is $\mathrm{supp}(w) \subseteq \mathrm{supp}(q)$. Let $\bar{\sigma}_{\min}(\boldsymbol{G})$ and $\sigma_{\max}(\boldsymbol{G})$ denote the minimum nonzero and maximum eigenvalues of $\boldsymbol{G}$, and let $\kappa_{\boldsymbol{G}} := \sigma_{\max}(\boldsymbol{G})/\bar{\sigma}_{\min}(\boldsymbol{G})$. If the vector target is represented by a WLS readout matrix $\boldsymbol{C}_{\boldsymbol{A}}$, define

$$R_{\boldsymbol{A}}^2 := \sigma_{\max}(\boldsymbol{C}_{\boldsymbol{A}}^\top \boldsymbol{G}^+ \boldsymbol{C}_{\boldsymbol{A}}), \qquad \|u\|_{2,\tilde{w}}^2 := \mathbb{E}_q[\tilde{w}(S)u(S)^2], \qquad \|u\|_{\infty,\tilde{w}}^2 := \sup_{S:q(S)>0} \tilde{w}(S)u(S)^2.$$

Assuming bounded utilities, finite L, stable nonzero spectrum of $\boldsymbol{G}$, and a ridge penalty on the order of $\bar{\sigma}_{\min}(\boldsymbol{G})$, the non-asymptotic WLS remainder bound has the form

$$\mathbb{E}_q\|\boldsymbol{r}_m\|_2^2 \le C_1 R_{\boldsymbol{A}}^2 L^2 \left[\frac{\log d_{\boldsymbol{z}}}{m^2}\left(1 + \frac{L \log d_{\boldsymbol{z}}}{m}\right)\|u\|_{2,\tilde{w}}^2 + e^{-C_2 m/L} d_{\boldsymbol{z}} m^2 \kappa_{\boldsymbol{G}}^2 \|u\|_{\infty,\tilde{w}}^2\right]. \tag{19}$$

For the centered Shapley WLS geometry, KernelSHAP has $L = O(n \log n)$ and LeverageSHAP has $L = O(n)$. This gives

$$\mathbb{E}_q\|\boldsymbol{r}_m\|_2^2 \le C_1 \|u\|_\infty^2 \begin{cases} \dfrac{n^2 \log^4 n}{m^2}\left(1 + \dfrac{n \log^2 n}{m}\right) & \text{KernelSHAP,} \\ \quad + n^3 m^2 \log^3 n \, e^{-C_2 m/(n \log n)}, & \\ \dfrac{n^2 \log^2 n}{m^2}\left(1 + \dfrac{n \log n}{m}\right) & \text{LeverageSHAP.} \\ \quad + n^4 m^2 e^{-C_2 m/n}, & \end{cases} \tag{20}$$

If the full-vector first-order variance has scale $O(n)$, then (20) makes the WLS remainder negligible at the relative-tolerance scales $m = \Omega(n \log^4 n/\epsilon^2)$ for KernelSHAP and $m = \Omega(n \log^2 n/\epsilon^2)$ for LeverageSHAP.

Appendix D. EASE Algorithm Details

This appendix gives the operational details behind Section 4. For a vector target, define

$$\boldsymbol{\rho}_{\boldsymbol{A}}(S) = (\rho_{\boldsymbol{a}_1}(S), \ldots, \rho_{\boldsymbol{a}_d}(S))^\top.$$

D.1. Budget Accounting and Fixed Conventions

All reported budgets count utility evaluations, not abstract sampling draws. For EASE, the reported budget is $m_{\text{init}} + m_{\text{est}}$, so pilot evaluations are charged. Cross-fitting only repartitions the final-stage observations and does not require additional calls to u. For any baseline that obtains multiple coalitions in one draw, each evaluated coalition is charged once. The stabilizing choices (B, ϵ, λ) and the pilot/final split are fixed before evaluation and are not tuned using the test errors plotted in the figures.

Algorithm 1: EASE with residual-aware sampling

Input: Partition $\mathscr{C} = \{C_1, \ldots, C_K\}$; pilot budget m_{init}; estimation budget m_{est}; fold partition $\{I_b\}_{b=1}^{B}$ of the final samples; target matrix $\boldsymbol{A}$; surrogate class $\mathcal{H}$; floor parameter $\epsilon \in (0, 1)$.

Output: Estimator $\hat{\boldsymbol{\tau}}_{\boldsymbol{A}}$.

For each cell, compute $M_{k,0} = |C_k|^{-1} \sum_{S \in C_k} \sum_j \rho_{\boldsymbol{a}_j}(S)^2$;

Set $q^{\text{init}}(S) \propto \sqrt{M_{k,0}}$ for $S \in C_k$;

Draw pilot samples $S_t^{\text{init}} \sim q^{\text{init}}$ and observe $Y_t^{\text{init}} = u(S_t^{\text{init}})$;

Fit an initial surrogate $\hat{\boldsymbol{h}}^{\text{init}}$ by the empirical version of (8);

Estimate $\hat{M}_k$ from pilot residuals as in (22);

Set $\tilde{q}(S) \propto \sqrt{\hat{M}_k}$ for $S \in C_k$ when the denominator is positive; otherwise set $\tilde{q} = q^{\text{init}}$;

Floor the sampling law: $q^{\text{final}} = (1 - \epsilon)\tilde{q} + \epsilon q^{\text{base}}$;

Draw final samples $S_t \sim q^{\text{final}}$ and observe $Y_t = u(S_t)$;

For each cross-fitting fold, fit $\hat{\boldsymbol{h}}^{(-b)}$ on the other folds and evaluate (7) on the held-out fold;

return the fold-averaged estimate $\hat{\boldsymbol{\tau}}_{\boldsymbol{A}}$;

D.2. Pilot Initialization and Cell Moments

For a law uniform within cells, write $p_k = \mathbb{P}_q(S \in C_k)$ for the cell mass and $q(S) = p_k/|C_k|$ for $S \in C_k$. For a fixed surrogate, minimizing the first-order risk over the cell masses gives $p_k \propto |C_k|\sqrt{M_k(\boldsymbol{h})}$, equivalently $q(S) \propto \sqrt{M_k(\boldsymbol{h})}$ within cell C_k, where M_k is defined in (10). Before fitting a surrogate, EASE uses the target-only proxy

$$M_{k,0} = \frac{1}{|C_k|} \sum_{S \in C_k} \sum_{j=1}^{d} \rho_{\boldsymbol{a}_j}(S)^2. \tag{21}$$

Thus the initialization law is

$$q^{\text{init}}(S) = \frac{\sqrt{M_{k,0}}}{\sum_{r=1}^{K} |C_r| \sqrt{M_{r,0}}}, \qquad S \in C_k,$$

with the obvious uniform fallback if all $M_{k,0}$ vanish. After pilot sampling, it estimates the residual moment by

$$\hat{M}_k = \frac{1}{\max\{N_k, 1\}} \sum_{t: S_t^{\text{init}} \in C_k} \sum_{j=1}^{d} \rho_{\boldsymbol{a}_j}(S_t^{\text{init}})^2 \{Y_t^{\text{init}} - \hat{h}_j^{\text{init}}(S_t^{\text{init}})\}^2, \tag{22}$$

where N_k is the number of pilot samples in cell C_k. Empty pilot cells receive $\hat{M}_k = 0$ before flooring. The unfloored plug-in law uses the same per-coalition form,

$$\tilde{q}(S) = \frac{\sqrt{\hat{M}_k}}{\sum_{r=1}^{K} |C_r| \sqrt{\hat{M}_r}}, \qquad S \in C_k,$$

when the denominator is positive, and otherwise reuses q^{init}. In all experiments, we set the pilot budget to one fifth of the total budget, $m_{\text{init}} = m/5$, and use the remaining evaluations for final-stage estimation.

D.3. Flooring the Learned Law

The unfloored law can assign too little probability to a cell whose pilot residual is underestimated. To avoid unstable inverse-probability weights, EASE uses

$$q^{\text{final}}(S) = (1-\epsilon)\tilde{q}(S) + \epsilon q^{\text{base}}(S), \qquad q^{\text{base}}(S) = \frac{1}{K|C_k|}, \quad S \in C_k. \tag{23}$$

The floor keeps every active cell sampled while preserving the residual-aware allocation when ϵ is small.

D.4. Shared Linear Surrogate

In the experiments, one may use a shared linear surrogate $h_{\boldsymbol{\beta}}(S) = \boldsymbol{x}(S)^\top \boldsymbol{\beta}$. For fixed q, define

$$\boldsymbol{\omega}_t = \boldsymbol{\rho}_{\boldsymbol{A}}(S_t)/q(S_t), \qquad a_t = \|\boldsymbol{\omega}_t\|_2^2, \qquad \boldsymbol{x}_t = \boldsymbol{x}(S_t).$$

The empirical analogue of $\mathscr{V}$ is

$$L_t(\boldsymbol{\beta}, \boldsymbol{\mu}) = \sum_{s=1}^{t} \|\boldsymbol{\omega}_s(Y_s - \boldsymbol{x}_s^\top \boldsymbol{\beta}) - \boldsymbol{\mu}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_2^2. \tag{24}$$

For fixed $\boldsymbol{\beta}$, the optimizing centering vector is the sample mean

$$\hat{\boldsymbol{\mu}}_t(\boldsymbol{\beta}) = \frac{1}{t} \sum_{s=1}^{t} \boldsymbol{\omega}_s(Y_s - \boldsymbol{x}_s^\top \boldsymbol{\beta}),$$

so optimizing over $\boldsymbol{\mu}$ is equivalent to subtracting the empirical mean of the weighted residuals before fitting $\boldsymbol{\beta}$. The minimizer can be updated from sufficient statistics

$$\boldsymbol{R}_t = \sum_{s=1}^{t} a_s \boldsymbol{x}_s \boldsymbol{x}_s^\top, \quad \boldsymbol{d}_t = \sum_{s=1}^{t} a_s \boldsymbol{x}_s Y_s, \quad \boldsymbol{U}_t = \sum_{s=1}^{t} \boldsymbol{\omega}_s \boldsymbol{x}_s^\top, \quad \boldsymbol{v}_t = \sum_{s=1}^{t} \boldsymbol{\omega}_s Y_s,$$

via

$$\hat{\boldsymbol{\beta}}_t = \left(\boldsymbol{R}_t - \frac{1}{t} \boldsymbol{U}_t^\top \boldsymbol{U}_t + \lambda \boldsymbol{I} \right)^{-1} \left(\boldsymbol{d}_t - \frac{1}{t} \boldsymbol{U}_t^\top \boldsymbol{v}_t \right). \tag{25}$$

Appendix E. Bundled Sampling

This section is not used by the main EASE experiments. It records how the same first-order view extends to coupled-coalition estimators used by some baselines. In these estimators, one sampling draw can return several utility evaluations, so the feasible weighting is a design object rather than being fixed by a single-coalition law.

E.1. General Bundles

In bundled sampling, one Monte Carlo draw returns

$$\boldsymbol{T} = (S_1, \ldots, S_r) \sim Q, \qquad u(\boldsymbol{T}) = (u(S_1), \ldots, u(S_r)).$$

Complement-pair sampling returns (S, S^c) [34]; edge-pair sampling returns $(S, S \cup \{i\})$ [16, 23]; permutation-path sampling returns the sequence of coalitions generated along a random permutation [1, 2, 22, 33]. In this notation, sampling lift, weighted sampling lift, permutation refinements, and paired WLS variants differ by the bundle law Q, the feasible weighting $\boldsymbol{\gamma}$, and the implicit or explicit surrogate class.

Definition 2 (Feasible bundle weighting) *For target $\tau_{\boldsymbol{a}}(u) = \sum_S \rho_{\boldsymbol{a}}(S)u(S)$ and bundle law Q, a vector function $\boldsymbol{\gamma}_{\boldsymbol{a},Q} : (2^{[n]})^r \to \mathbb{R}^r$ is feasible if*

$$\tau_{\boldsymbol{a}}(u) = \mathbb{E}_Q[\boldsymbol{\gamma}_{\boldsymbol{a},Q}(\boldsymbol{T})^\top u(\boldsymbol{T})]$$

for every utility u.

Equivalently, feasibility requires

$$\rho_{\boldsymbol{a}}(S) = \mathbb{E}_Q\left[\sum_{\ell=1}^{r} \gamma_{\boldsymbol{a},Q,\ell}(\boldsymbol{T})\mathbf{1}\{S_\ell = S\}\right], \qquad S \subseteq [n]. \tag{26}$$

For single-coalition sampling, this fixes $\gamma_{\boldsymbol{a},q}(S) = \rho_{\boldsymbol{a}}(S)/q(S)$. For $r \geq 2$, it is usually an underdetermined linear system, so the weighting becomes part of estimator design.

E.2. Complement-Pair Example

Let $\boldsymbol{T} = (S, S^c)$ with $S \sim q$. A feasible weighting must satisfy

$$q(S)\gamma_1(S) + q(S^c)\gamma_2(S^c) = \rho_{\boldsymbol{a}}(S), \qquad S \subseteq [n].$$

If the target has complement antisymmetry, $\rho_{\boldsymbol{a}}(S^c) = -\rho_{\boldsymbol{a}}(S)$, the canonical paired weighting is

$$\gamma_1(S, S^c) = \frac{\rho_{\boldsymbol{a}}(S)}{q(S) + q(S^c)}, \qquad \gamma_2(S, S^c) = \frac{\rho_{\boldsymbol{a}}(S^c)}{q(S) + q(S^c)}. \tag{27}$$

This is the Rao–Blackwellized weighting for a fixed complement-pair law and surrogate in the symmetric setting.

E.3. Bundle First-Order Risk

For a feasible bundle weighting $\boldsymbol{\gamma}$ and surrogate h, define

$$\psi_{\boldsymbol{\gamma},h}(\boldsymbol{T}; u) = \boldsymbol{\gamma}(\boldsymbol{T})^\top\{u(\boldsymbol{T}) - h(\boldsymbol{T})\} + \tau_{\boldsymbol{a}}(h) - \tau_{\boldsymbol{a}}(u).$$

The bundle analogue of first-order risk is

$$V_Q(\boldsymbol{\gamma}, h) = \mathrm{Var}_Q\left[\boldsymbol{\gamma}(\boldsymbol{T})^\top\{u(\boldsymbol{T}) - h(\boldsymbol{T})\}\right].$$

Thus bundled designs require choosing not only a bundle law Q and a surrogate h, but also a feasible weighting $\boldsymbol{\gamma}$.

Appendix F. Experimental Details and Additional Results

F.1. SOU Game Construction

The experiments use structured sum-of-unanimity games

$$u(S) = \sum_{\tilde{S} \in \mathscr{S}} \theta_{\tilde{S}} \mathbf{1}\{\tilde{S} \subseteq S\}.$$

Each component contributes only when all players in $\tilde{S}$ are present. This benchmark is useful because probabilistic values can be computed exactly by linearity over unanimity-game components, so relative squared error is measured against ground truth.

We use $n = 40$ players. The component multiset is $\mathscr{S} = \mathscr{S}_{\text{low}} \cup \mathscr{S}_{\text{high}}$, where $\mathscr{S}_{\text{low}} = \{\tilde{S} \subseteq [n] : 1 \leq |\tilde{S}| \leq 2\}$ contains all singleton and pairwise coalitions, and $\mathscr{S}_{\text{high}}$ contains n^2 randomly sampled coalitions of size at least three. For $\eta \in (0, 1)$ and scale σ^2, coefficients are drawn independently as

$$\theta_{\tilde{S}} \sim N\left(0, \frac{\eta\sigma^2}{|\mathscr{S}_{\text{low}}|}\right), \quad \tilde{S} \in \mathscr{S}_{\text{low}}, \qquad \theta_{\tilde{S}} \sim N\left(0, \frac{(1-\eta)\sigma^2}{|\mathscr{S}_{\text{high}}|}\right), \quad \tilde{S} \in \mathscr{S}_{\text{high}}.$$

Under this scaling, $\eta\sigma^2$ is the total coefficient variance assigned to low-order components and $(1-\eta)\sigma^2$ to high-order components. Larger η makes the utility easier for low-order surrogates to approximate, so varying η gives a controlled range of structural complexity. The main text uses the fixed setting $\eta = 0.25$; Figures 3 and 4 report additional settings $\eta = 0.5$ and $\eta = 0.75$.

F.2. EASE Variants

The AUCC benchmark uses two EASE surrogate classes. EASE-FO (First-Order) uses player indicators $\mathbf{1}\{i \in S\}$, augmented with an intercept and two size features $\log(1 + |S|)$ and $(|S|/n)^2$. EASE-SP (Size-Player) uses the richer player-by-size indicators $\mathbf{1}\{|S| = s, i \in S\}$.

F.3. Matched Surrogate Classes

The matched comparisons in Figure 1 hold the working class fixed within each panel:

$$\begin{aligned}
&\text{RegressionMSR match:} && \operatorname{span}\{\mathbf{1}(i \in S) : i \in [n]\},\\
&\text{OFA match:} && \operatorname{span}\{\mathbf{1}(|S| = s, i \in S) : s = 0, \ldots, n,\ i \in [n]\},\\
&\text{PolySHAP match:} && \operatorname{span}\{\mathbf{1}(\tilde{S} \subseteq S) : |\tilde{S}| \leq 2\}.
\end{aligned}$$

These comparisons isolate the effect of the efficiency-aware loss and residual-aware sampling law.

F.4. Additional Matched Comparisons

Figure 3 repeats the one-on-one matched comparisons for $\eta = 0.5$ and $\eta = 0.75$. Together with the $\eta = 0.25$ main-text panel, these results show the full SOU benchmark trend: across all three matched comparisons and all three η settings, EASE has lower Shapley-value relative squared error than the corresponding baseline at every measured budget. Increasing η puts more coefficient variance in low-order unanimity terms, so all matched low-order surrogate methods improve; the degree-two PolySHAP-matched class benefits especially at $\eta = 0.75$, where main effects and pairwise interactions explain more of the utility. The gap against RegressionMSR narrows as the first-order player surrogate becomes more adequate, while EASE remains strong against OFA and PolySHAP because the residual-aware sampling law still targets the variation left by the matched surrogate.

F.5. AUCC Metric and Baselines

For budgets $\mathcal{B} = \{50, 100, \ldots, 5000\}$, AUCC is

$$\text{AUCC} = \frac{1}{|\mathcal{B}|} \sum_{m \in \mathcal{B}} \frac{\|\hat{\boldsymbol{\phi}}_m - \boldsymbol{\phi}\|_2^2}{\|\boldsymbol{\phi}\|_2^2}.$$

The broad benchmark groups methods by how they are used in the experiments; Table 1 gives the theoretical first-order taxonomy. For example, OFA appears below with normalization/lift-style baselines because of its empirical role, while Table 1 records the cellwise class induced by its normalization.

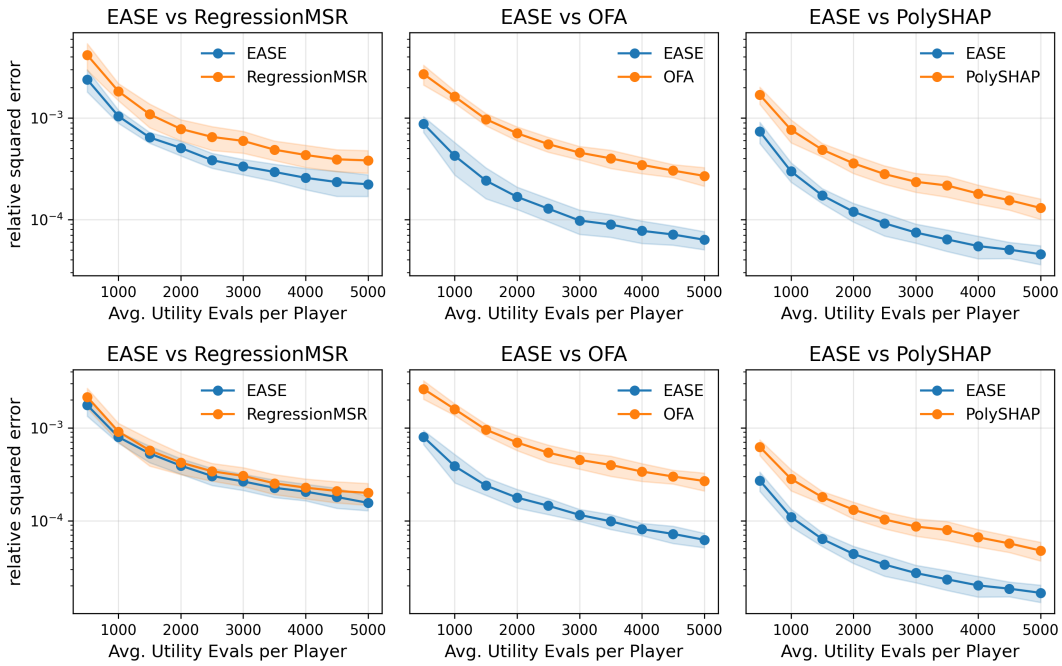


Figure 3: Matched comparisons on SOU games for $\eta = 0.5$ (top) and $\eta = 0.75$ (bottom). Each panel keeps the same working surrogate class for EASE and the baseline, so improvements isolate the efficiency-aware loss and residual-aware sampling design. Curves show mean relative squared error, with shaded bands indicating one standard deviation over 10 runs.

- normalization, weighted-average, and lift estimators: OFA [18], sampling lift and weighted sampling lift [16, 23], and SHAP-IQ [9];
- WLS and regression estimators: GELS/WGELS [19], KernelSHAP [5, 21], Regression-MSR [32], and PolySHAP [10];
- additional baselines: permutation and weighted variants [1], complementary-contribution sampling [34], group-testing data valuation [13], and AME [20].

Methods restricted to particular probabilistic values are evaluated only where applicable.

F.6. Additional Benchmark Results for SOU Games

Figure 4 reports additional AUCC benchmarks on SOU games with player counts $n \in \{40, 80, 160\}$ and structural parameters $\eta \in \{0.25, 0.5, 0.75\}$. Across these settings, EASE-FO and EASE-SP continue to outperform most competing estimators, showing that efficiency-aware surrogate fitting and residual-aware sampling remain effective as both the game size and low-order signal strength vary. The relative performance of the two EASE variants is more nuanced. EASE-SP often performs slightly better for smaller games, where its more expressive player-by-size surrogate can be fitted reliably. However, it can underperform EASE-FO in some larger-n settings, especially for $\mathrm{WB}(0.5)$. This degradation is consistent with the complexity of the EASE-SP surrogate: the

player-by-size feature class has order n^2 dimensions, and therefore, under small evaluation budgets m with relatively large n, fitting this richer surrogate can become unstable.

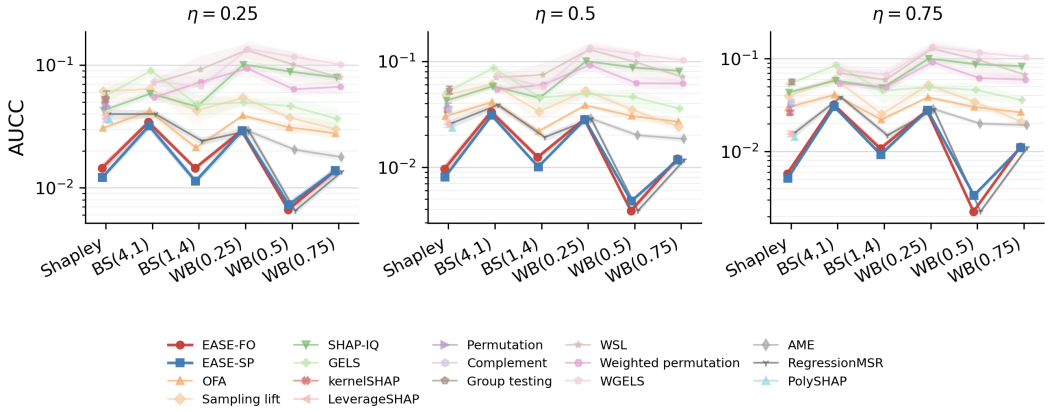
F.7. Benchmark Results for Real-World Feature-Attribution Tasks

We further benchmark EASE against the baseline methods on five real-world feature-attribution tasks: ViT on ImageNet [6], CIFAR-10 [15], BreastCancer [29], NHANES I [25], and Communities & Crime [27]. In these games, the players correspond to either image patch groups or tabular input features, and the utility of a coalition is defined as the prediction model output when only the corresponding features are observed. Table 2 summarizes the dataset, player count, and prediction model used for each benchmark. For each dataset and probabilistic-value target, AUCC is computed from 10 budget checkpoints up to a final budget of $m = 40n$ utility evaluations.

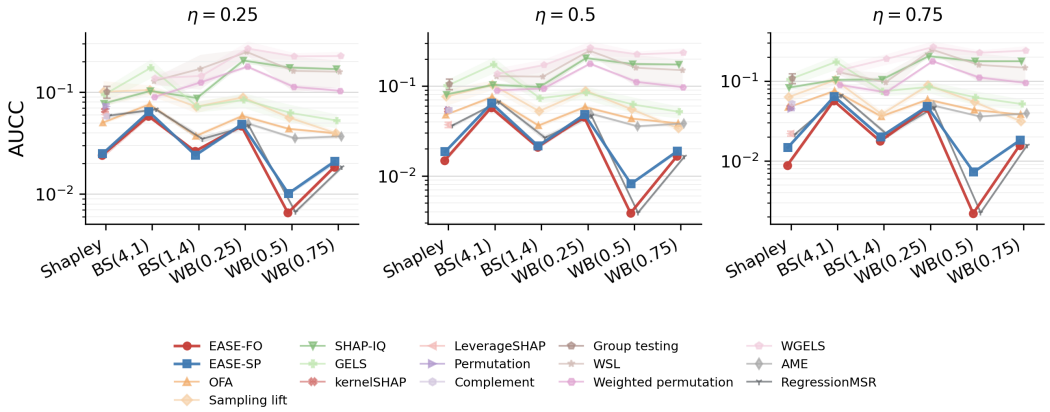
Table 2: Real-world feature-attribution benchmark settings. The player count n is the number of image patch groups or tabular features used in the coalition game. RF denotes random forest.

Dataset	n	Model
ViT [6]	16	pretrained ViT
CIFAR-10 [15]	16	fine-tuned ViT
BreastCancer [29]	30	RF classifier
NHANES I [25]	79	RF
Communities & Crime [27]	101	RF regressor

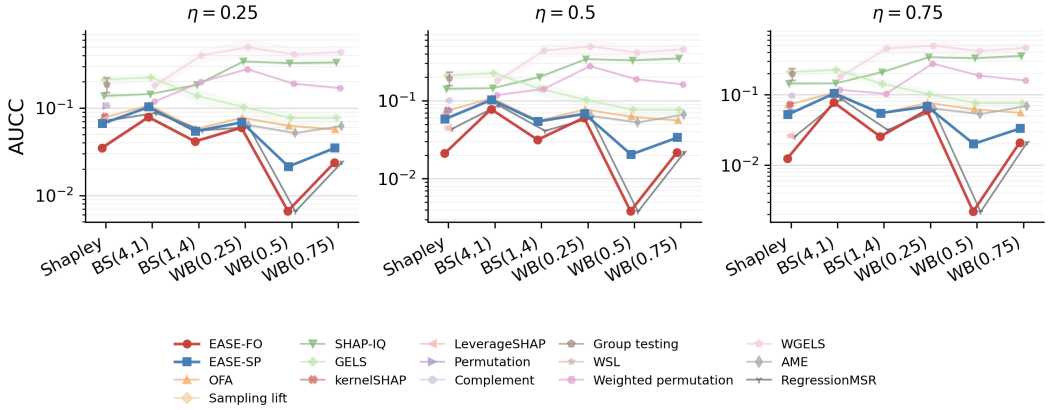
Figure 5 reports the AUCC results. Overall, the real-world benchmarks show the same qualitative pattern as the SOU experiments, despite involving less controlled utility functions. Across the five tasks and six probabilistic-value targets, EASE-FO is consistently among the strongest methods and typically attains the lowest AUCC. This suggests that the efficiency-aware first-order surrogate remains effective beyond synthetic SOU games. The performance of EASE-SP is less stable on these real-world tasks. Its richer player-by-size surrogate can be beneficial in lower-dimensional settings, but because this surrogate class has order-n^2 dimensions, it becomes harder to estimate accurately under the relatively small budget of $40n$ utility evaluations. This effect is most apparent on the tabular games with larger number of players, where the limited evaluation budget favors the more parsimonious first-order surrogate used by EASE-FO.



(a) $n = 40$



(b) $n = 80$



(c) $n = 160$

Figure 4: Additional SOU benchmark AUCC results. Rows use $n = 40$, 80, and 160; columns use $\eta = 0.25$, 0.5, and 0.75. Lower is better.

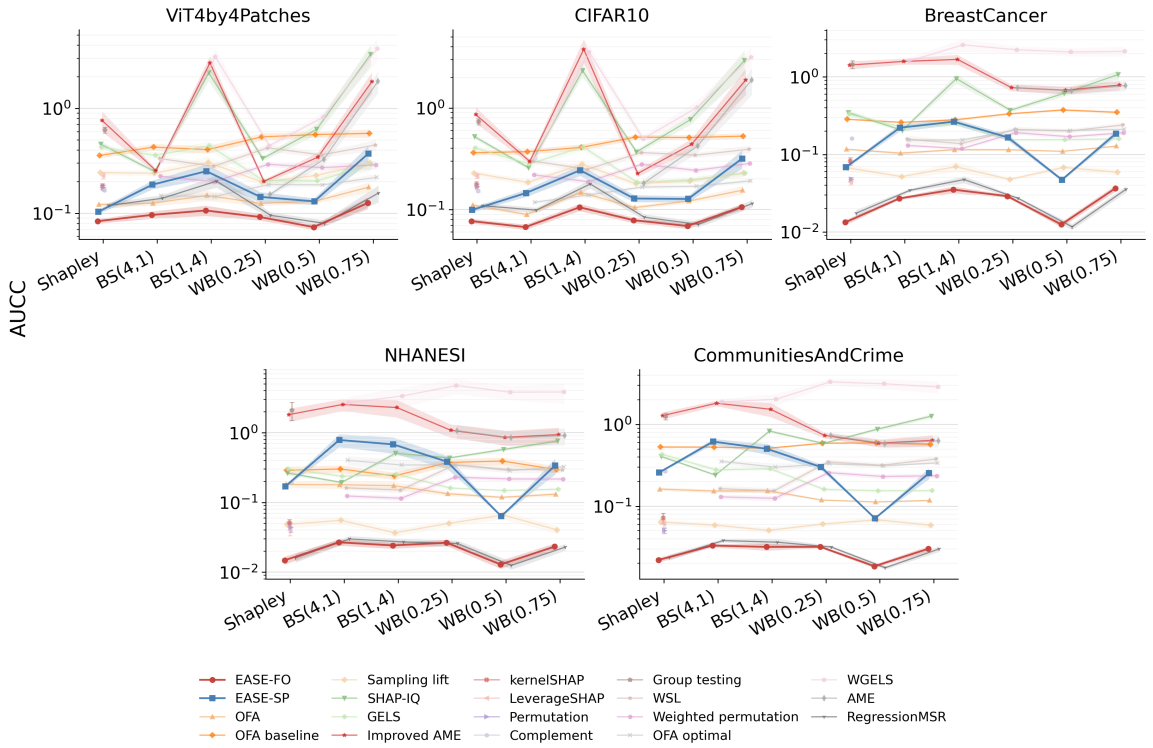


Figure 5: Real-world feature-attribution AUCC benchmark. Each panel is one dataset, and the x-axis lists the target probabilistic value. AUCC averages 10 budget checkpoints up to $m = 40n$ utility evaluations. Lower is better.