University of St.Gallen
Institute of Insurance Economics

Limited information and its Impact on a Policyholder's optimal choice on Deductibles

Jan-Christian fey
Hato Schmeiser
Florian Schreiber

Working Papers on Risk Management and Insurance No. 257

Edited by Hato Schmeiser
Chair for Risk Management and Insurance

August 2022

Limited Information and its Impact on a Policyholder's Optimal Choice on Deductibles*

Jan-Christian Fey[†1], Hato Schmeiser[1], and Florian Schreiber[2]

[1]University of St. Gallen

[2]Lucerne University of Applied Science and Arts

August 29, 2022

Abstract

When determining the optimal deductible level for an insurance policy, a policyholder faces two sources of uncertainty. First, uncertainty arises from the randomness of future losses. Second, the opacity of the functional forms of the policyholder's loss distribution and utility function also contributes to uncertainty. While the academic literature focuses on the former, we additionally include limited information on these functional forms in our model setting to reflect real-world decision-making. That is, we draw on an expected utility framework and analyze the relationship between optimal deductible levels under limited and full information. We also derive several decision rules under limited information in order to approximate the optimal deductible level under full information. To support real-world decision-making, these rules could be easily implemented in an online decision aid.

Keywords: Optimal deductible choice, limited information, policyholder decision rules, welfare economics

JEL codes: D81, D83, G22

*We gratefully acknowledge research funding from the Basic Research Fund (GFF) of the University of St. Gallen (project number 2050181). Martin Boyer, Martin Brown, Martin Eling, Helmut Gründl, Justin Sydnor, and Kar Man Tan provided very helpful feedback. Moreover, we appreciate helpful comments from participants of the 2022 Deutscher Verein für Versicherungswissenschaft e.V. (DVfVW) Annual Meeting and the 2022 American Risk and Insurance Association (ARIA) Annual Meeting.

†Corresponding author: Jan-Christian Fey, jan-christian.fey@unisg.ch

1 Introduction

In today's insurance markets, deductibles are a predominant feature of insurance contracts. For a risk-averse expected utility maximizer, several authors have shown that deductible insurance policies are the optimal contract type under typical market conditions (cf., e.g., Arrow, 1963; Raviv, 1979; Gollier, 2013). In practice, determining the optimal deductible level is one of the biggest challenges policyholders face.

In theory, policyholders' deductible choices depend on their loss and individual wealth distributions, their risk preferences, and the premium offered by insurance providers. With full knowledge of the loss distribution and the utility function, the optimal deductible level can be determined by optimizing the expected utility with respect to the available deductible levels. In reality, however, policyholders have limited information about their loss distribution and preferences, making it difficult for them to determine their optimal deductible levels.

Empirical research on deductible choices suggests that a substantial fraction of policyholders choose non-optimal deductibles, which can lead to significant welfare losses. For example, Feldstein (1973) has estimated that the welfare losses suffered by U.S. families from overinsurance against healthcare expenditures range from \$2.4 to \$3.7 billion per year. The recent study of Bhargava et al. (2017) also shows that the majority of employees enrolled in a Fortune 100 company health plan choose policies with deductible levels that are strictly dominated by other plan options. By switching to a non-dominated plan, those with a non-optimal deductible could have saved an average of \$375 per year, a significant portion of the net annual income for many workers. Moreover, a particularly worrisome finding of their analysis is that the share of individuals opting for dominated policies is especially high among low-income individuals. Similar results are also found for other insurance products and countries (cf., e.g., Sydnor, 2010; Van Winssen et al., 2015, 2016).

Given this clear empirical evidence and the absolute level of insurance premiums paid by individuals, it is crucial to provide policyholders with appropriate tools for choosing the optimal deductible level. Since these decision aids must rely on limited information about important decision criteria, one must adapt the usual model frameworks under full information to derive appropriate decision rules. To date, a large strand of literature has used an expected utility framework to study a policyholder's optimal deductible choice (cf., e.g., Mossin, 1968; Moffet, 1977; Schlesinger, 1981, 2013, 1985; Doherty and Schlesinger, 1983). A key assumption of all these analyses is that the policyholder fully knows the functional forms of her loss distribution as well as her utility function. In these studies, uncertainty arises only from the randomness of future losses. However, if we consider decision making in the real world, there is a second level of uncertainty that has been largely ignored in the literature – that is, uncertainty about the functional forms of the above-mentioned decision making factors.

Although existing models that assume complete information provide important information about cause-effect relationships, they are simply not applicable by customers because important input data is missing. Therefore, the goal of our paper is to develop a model framework that helps customers make the best possible deductible choice given their limited information. To do so, we model this type of limited information and integrate it into an expected utility framework. We consider observed loss realizations, available deductible policies, and information obtained from utility elicitation methods such as the probability equivalence (PE) method as sources of information for approximating the functional forms. This model framework allows us to examine how optimal deductible choices under limited and full information are related, as well as to derive valuable insights for insurance decision making in the real world. In this way, we apply important concepts of heuristic decision making and decision analysis to the theory of optimal deductible choice. More concretely, we follow the four steps of decision analysis in the the sense of Keeney (1982) in order to derive suitable approximations of the optimal deductible choice. By doing so, we extend the vast literature on decisions analysis applications in operations research (cf., e.g., Corner and Kirkwood, 1991) by an application in risk management.

In a first step, the question we seek to address is how to derive a suitable approximation of the loss distribution. Using the observed loss realizations, a policyholder can derive the empirical distribution function. In addition, there are ways to obtain information on the distribution from the policy mix offered. We show one way to extract this information to obtain another approximation of the loss distribution. Based on these two approximations, we then discuss how they can be combined into one approximation. Our theoretical results as well as a numerical illustration of rule-based decision making demonstrate that policyholders should mainly rely on the approximation derived from the available deductible policies when the number of observed loss realizations is small and the available deductibles levels cover a significant portion of the loss range. Since most policyholders probably disregard distributional information from the policy mix offered, our results suggest that there is much scope to improve real-world deductible choices by leveraging this source of information. In a numerical example, we illustrate this improvement by comparing expected costs of choosing a wrong deductible for different decision aids. That is, we calculate the economic welfare gains of policyholders that would result from a deductible choice based on a decision rule which fully leverages all available distributional information.

In a second step, we also consider information on risk preferences modeled by von Neumann-

Morgenstern utility functions. This information is obtained using the probability equivalence (PE) method. For this method, we discuss how much and in what way preference information should be acquired. More concretely, we present a stepwise survey design for assessing utility functions. This procedure has two major advantages over commonly used survey designs (cf., e.g., Murray, 1972; Hershey et al., 1982; Hershey and Schoemaker, 1985; Bleichrodt et al., 2001). First, it leverages information on the distribution of risk preferences in the overall population. Second, it covers the curvature of a utility function more accurately. Moreover, we analyze the consequences that possible biases in the PE method may have on recommendations for optimal deductible levels. After deriving the theoretical implications of these biases, we introduce a bias modeled by Bleichrodt et al. (2001).

Given the above-mentioned information, we discuss several decision rules for estimating the optimal deductible choice under limited information. As a first step, we define desirable properties for the decision rules and categorize them according to these properties. Next, we compare the performances of the decision rules using a numerical illustration of the U.S. homeowners insurance market. In this analysis, we account for the large heterogeneity in policyholders' attitudes toward risk by comparing the decision rules across different risk types (cf., e.g., Cohen and Einav, 2007; Ericson and Starc, 2012). We also vary the extent to which distributional and preference information is available to assess in which situations a particular decision aid performs better than the other available rules. As a result, we find that decision rules that use both preference information and information about the two types of loss distribution approximations produce, by far, the lowest expected costs of choosing an inappropriate deductible. However, these decision rules are susceptible to biases in the elicitation process.

To support real-world decision making, our model as well as the corresponding decision rules could be easily implemented online and made available to policyholders so that they can receive a suggestion for the optimal deductible level based on the information they provide and simple preference tests.

The remainder of this paper is organized as follows. In the next section, we introduce our model framework and discuss different types of limited information in detail. The third section addresses how to approximate the unknown loss distribution using the available information sources. In the fourth section, we analyze the impact of preference uncertainty on the optimal deductible choice under limited information and discuss how to approximate a policyholder's utility function. In the fifth section, we then examine decision rules for choosing the optimal deductible level under limited information. The sixth section highlights the theoretical results in a numerical illustration of a loss model motivated by the U.S. homeowners insurance market.

Finally, in the last section, we summarize our findings, discuss the policy implications, and draw our conclusion.

2 Model Setting

In our model, we consider observed loss realizations, the policy mix offered, and a finite number of functional values of the policyholder's utility function as sources of information.

For the analysis, we further assume that there is no adverse selection or moral hazard; that is, the probability distribution for losses is identical for all available deductible choices. An overview of insurance markets for which this assumption holds can be found in Cohen and Siegelman (2010). The loss severity is modeled by the random variable L, whose cumulative distribution function is given by

$$F(x) = \mathbb{P}(L \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ p & \text{if } x = 0 \\ p + \int_0^x f(z)\, dz & \text{if } 0 < x < N \\ 1 & \text{if } x \geq N \end{cases},$$

where f is a continuous non-negative function, $p > 0$ and $N > 0$.[1] For the rest of this paper, we assume that the first two central moments of this distribution, i.e., mean $\mu = \mathbb{E}[L]$ and variance $\sigma^2 = \mathrm{Var}[L]$, are information attainable by policyholders.

To protect against a possible loss, deductible insurance is available at deducible levels $0 < D_1 < \cdots < D_K < N$. In addition to these deductible levels, the policyholder can always choose a deductible level N, which is equivalent to purchasing no insurance coverage. The premiums for an insurance policy with deductible level D are determined by the function $R(D) = (1 + \lambda)\, C(D)$ for a loading $\lambda > 0$, where

$$C(D) = \int_D^N (z - D)\, f(z)\, dz$$

is the actuarial fair value of the corresponding policy.[2]

The individual's wealth is given by a deterministic variable W satisfying the condition $W > N$. Thus, the final wealth of the policyholder for a chosen deductible level D is given by the random

[1] This form of probability distribution is derived from Schlesinger (1981). Compared to Schlesinger (1981), however, the probability distribution has no mass $q > 0$ at $x = N$ for computational reasons.

[2] Please note that we refrain from discounting, since the current risk-free rate is almost equal to 0%.

variable

$$Y_D(L) = \begin{cases} W - R(D) & \text{if } L = 0 \\ W - R(D) - L & \text{if } 0 < L \leq D \\ W - R(D) - D & \text{if } L > D \end{cases}.$$

The policyholder's preferences are modeled by a utility function u, assuming that u is everywhere twice differentiable with $u' > 0$ and $u'' \leq 0$. We also assume that $u(W - N) = 0$ and $u(W) = 1$. Given this utility function, the expected utility from choosing an insurance policy with deductible level D is $\mathbb{E}\left[u\left(Y_D\left(L\right)\right)\right]$. The policyholder's objective is to choose the deductible level D_K^* that maximizes the expected utility $\mathbb{E}\left[u\left(Y_D\left(L\right)\right)\right]$; that is,

$$D_K^* = \underset{D \in \{D_1, \ldots, D_K, N\}}{\arg\max} \; \mathbb{E}\left[u\left(Y_D\left(L\right)\right)\right]. \tag{1}$$

Next, we introduce limited information into the model framework. We assume that the insurance buyer does not know the exact functional form of the cumulative distribution function F. For this reason, she needs to derive an approximation of this distribution. To this end, she has two sources of information at her disposal. First, she observes n independent and identically distributed realizations of the loss variable L, which we denote by $L_1, \ldots, L_n$.[3] Moreover, she can infer information about the functional form of F from the insurance premiums. With respect to premiums, we assume that the potential policyholder can also determine the loading λ, so that she can calculate the actuarial fair value $C(D)$.[4]

Regarding the policyholder's preferences, we introduce limited information about the shape of the utility function by assuming that the policyholder only knows m functional values of the utility function u in the interval $(W - N, W)$. That is, she knows the values $u(x_1), \ldots, u(x_m)$ for $W - N < x_1 < \cdots < x_m < W$. Moreover, we also assume that she knows that $u(W - N) = 0$ and $u(W) = 1$. The motivation behind this type of information about the policyholder's preferences is as follows (cf., e.g., Murray, 1972): The policyholder answers a survey in which she is presented with hypothetical decision situations. In each situation, she can choose between two different actions a_1 and a_2. Action a_1 leads to $x \in (W - N, W)$ with probability 1, while action a_2 leads to $W - N$ or W with probabilities $1 - p$ and p (cf. Figure 1). The probability p varies until the

[3] Insurance associations and regulators typically provide annual data regarding incurred losses for each line of business which is publicly available.

[4] Analogous to incurred losses, insurance associations typically provide further data, such as the (aggregated) expense ratio of the industry. This data, in turn, could be tapped by consumer protection organizations or comparison portals and made transparently available to interested parties.

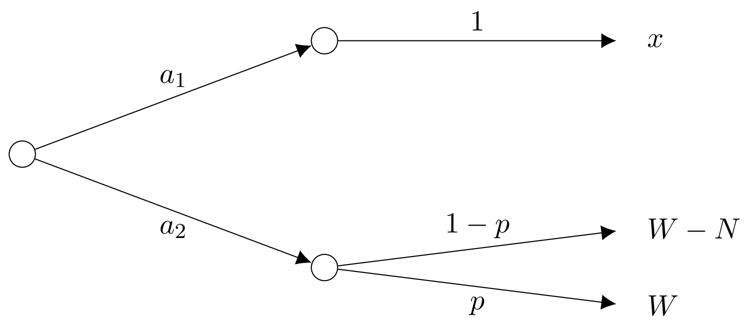


Figure 1: Illustration of the hypothetical decision situations faced by the policyholder

policyholder is indifferent between the actions a_1 and a_2. Thus, we obtain

$$u(x) = (1-p^*)u(W-N) + p^*u(W) = p^*$$

for the indifference probability p^*. In the survey, the policyholder identifies the indifference probabilities $p_1^*, \dots, p_m^*$ for the values $W - N < x_1 < \cdots < x_m < W$, so that she knows the functional values $u(x_1), \dots, u(x_m)$.[5]

This elicitation method is called the probability equivalence (PE) method, which together with the certainty equivalence (CE) method (cf., e.g., Hershey and Schoemaker, 1985; Bleichrodt et al., 2001) is one of the most commonly used elicitation methods.[6] Theoretically, all of the above elicitation processes should result in the same utility values, so it should not matter which method is used. However, there is ample experimental evidence that the different methods lead to different utility functions (cf., e.g., Hershey et al., 1982; Hershey and Schoemaker, 1985; Bleichrodt et al., 2001). For example, Hershey and Schoemaker (1985) find experimental evidence that the PE method leads to more risk-averse preferences than the CE method. The authors cite reframing effects, anchoring, salience effects, strategic misrepresentation, regret or rejoice influences, and endowment effects as possible explanations for their findings. Thus, it could be the case that certain methods lead to a systematic bias in the utility function, which in turn could affect the estimation of a policyholder's optimal deductible D_K^*.

To examine the effects of this bias, suppose that the utility function resulting from an elicitation process is denoted by $u^{(b)}$ and that the relationship between u and $u^{(b)}$ is modeled by

$$u(x) = s(u^{(b)}(x)),$$

[5] In general, it is possible to account for random noise in the elicitation process as discussed, for example, in Chapter 5 in Hershey and Schoemaker (1985). However, we choose not to do so in order to keep our elicitation process simple. When incorporating random noise into the model, one way to account for this noise is to fit a non-decreasing function to the measured utility values. For evenly distributed measurement errors with mean zero, such approximations improve with an increasing number of functional values m.

[6] Please note that in addition to these commonly used methods, other procedures also exist, such as the tradeoff method (Wakker and Deneffe, 1996).

where s is a twice differentiable function, which we call the correction function. If the inverse function s^{-1} exists, we call this function a distortion function, since it explains the distortions of $u^{(b)}$ from u by the relationship

$$u^{(b)}(x) = s^{-1}(u(x)).$$

Then, the optimal deductible under $u^{(b)}$ is given by

$$D_K^{(b)} = \underset{D \in \{D_1,\ldots,D_K,N\}}{\arg\max} \ \mathbb{E}\left[u^{(b)}\left(Y_D\left(L\right)\right)\right]. \tag{2}$$

For the relationship of $D_K^{(b)}$ to the optimal deductible D_K^*, we get the following results:

Theorem 2.1. *Assume that $s' > 0$. Then, the following relationships between D_K^* and $D_K^{(b)}$ hold:*

(i) If $s'' < 0$, $D_K^* \leq D_K^{(b)}$.

(ii) If $s'' > 0$, $D_K^* \geq D_K^{(b)}$.[7]

Unfortunately, the results of Theorem 2.1 only hold if we make the rather strict assumptions $s'' < 0$ or $s'' > 0$. If these assumptions are violated, no general conclusions can be drawn about the relationship between D_K^* and $D_K^{(b)}$.[8]

3 Deriving an Approximation of the Loss Distribution

Based on the observed loss realizations $L_1, \ldots, L_n$, one can derive the empirical distribution function

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(L_i).$$

According to the Glivenko-Cantelli Theorem (cf., e.g., Theorem 5.23 in Klenke (2013)),

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} \left|\hat{F}_n(x) - F(x)\right| = 0$$

$\mathbb{P}$-almost surely. Thus, the empirical distribution function converges to the cumulative distribution function F of the loss random variable L as n goes toward ∞.

[7]Please note that the proofs of the theorems and lemmas can be found in Appendix A.

[8]In our numerical illustration of rule-based decision making presented in Section 6, we assume a specific functional form for s introduced by Bleichrodt et al. (2001), which is discussed in Appendix C. If one additionally uses a specific correction for the CE method, which is beyond the focus of our paper, the differences between the PE and the CE method disappear completely.

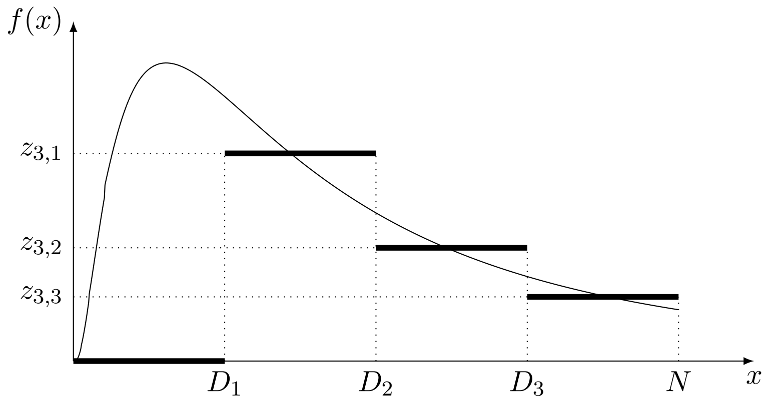


Figure 2: Illustration of the density function $\hat{\hat{f}}_K^{ded}$ for $K = 3$

Next, we derive an approximation of the loss distribution from the premiums for the various available deductible choices $D_1, \ldots, D_K$, which we denote by $\hat{\hat{F}}_K^{ded}$. This approximation is characterized by a probability density function $\hat{\hat{f}}_K^{ded}$ of the form

$$\hat{\hat{f}}_K^{ded}(x) = \mathbb{1}_{(D_1,D_2]}(x) z_{K,1} + ... + \mathbb{1}_{(D_K,N]}(x) z_{K,K},$$

which is illustrated in Figure 2 for $K = 3$, and

$$\hat{\hat{F}}_k^{ded}(0) = \hat{\hat{p}}_K = 1 - \int_0^N \hat{\hat{f}}_K^{ded}(z)dz = 1 - \sum_{i=1}^{K-1} z_{K,i}(D_{i+1} - D_i) - z_{K,K}(N - D_K). \quad (3)$$

Thus, we need to find appropriate values $\hat{\hat{p}}_K, z_{K,1}, \ldots, z_{K,K}$ that lead to the best possible approximation of the cumulative distribution function F. To do this, we exploit the fact that the policyholder can derive the actuarial fair values $C(D_1), \ldots, C(D_K)$ from the premiums because she can estimate the loading λ.[9] Now we want to ensure that the probability density approximation $\hat{\hat{f}}_K^{ded}$ leads to the same actuarial fair values. That is, $z_{K,1}, \ldots, z_{K,K}$ must satisfy the equations

$$C(D_k) = \sum_{i=k}^{K-1} z_{K,i} \int_{D_i}^{D_{i+1}} (y - D_k)\, dy + z_{K,K} \int_{D_K}^{N} (y - D_k)\, dy \quad (4)$$

for $k \in \{1, \ldots, K\}$. Thus, from (3) and (4) we obtain a system of linear equations with $K + 1$ equations and $K + 1$ unknown variables. In matrix notation, we can represent this system of

[9] As noted in Section 2, insurance associations typically provide various industry-level financial metrics so that actuarially fair premiums can be roughly estimated on this basis.

linear equations as follows:

$$\begin{pmatrix} 1 & D_2 - D_1 & \cdots & \cdots & N - D_K \\ 0 & M_{2,2} & \cdots & \cdots & M_{2(K+1)} \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & 0 & M_{KK} & M_{K(K+1)} \\ 0 & \cdots & \cdots & 0 & M_{(K+1)(K+1)} \end{pmatrix} \begin{pmatrix} \hat{\hat{p}}_K \\ z_{K,1} \\ \vdots \\ z_{K,K} \end{pmatrix} = \begin{pmatrix} 1 \\ C(D_1) \\ \vdots \\ C(D_K) \end{pmatrix},$$

where

$$M_{ij} = \begin{cases} 0 & \text{if } i > j \\ \int_{D_{j-1}}^{D_j} (y - D_i) dy & \text{if } i \leq j, j < K+1 \\ \int_{D_K}^{N} (y - D_i) dy & \text{if } i \leq j, j = K+1 \end{cases}$$

for $i, j \in \{2, \dots, K+1\}$. Thus, we see that the matrix representing our system of linear equations has rank $K+1$, so there is a unique solution for the system in (3) and (4).

Next, we analyze under what conditions the estimated distribution function $\hat{\hat{F}}_K^{ded}$ converges to the cumulative distribution function F when K goes to ∞. For this purpose, we denote the partition

$$0 = D_0 < D_1 < \cdots < D_K < D_{K+1} = N,$$

where $D_1, \dots, D_K$ are the available deductible choices, by $\bar{D}_K$. The norm of this partition is defined by the function

$$\delta(\bar{D}_K) := \max_{i \in \{0, \dots, K\}} |D_{i+1} - D_i|,$$

which measures the maximum distance between adjacent points in the set $\bar{D}_K$. We can now derive a theorem that states under what conditions the estimated distribution function $\hat{\hat{F}}_K^{ded}$ converges to the cumulative distribution function F.

Theorem 3.1. *Let $(D_K)_{K\in\mathbb{N}}$ be a sequence of deductible choices and denote by $\bar{D}_K$ the partition of the first K deductible choices of this sequence. Moreover, suppose that $\lim_{K\to\infty} \delta(\bar{D}_K) = 0$. Then, we obtain*

$$\limsup_{K\to\infty} \sup_{x\in\mathbb{R}} \left| \hat{\hat{F}}_K^{ded}(x) - F(x) \right| = 0. \tag{5}$$

According to Theorem 3.1, the convergence of $\hat{\hat{F}}_K^{ded}$ to F for $K \to \infty$ is guaranteed if the sequence of deductible choices $(D_K)_{K\in\mathbb{N}}$ adequately covers the entire interval $(0, N)$, which is achieved by assuming that $\lim_{K\uparrow\infty} \delta(\bar{D}_K) = 0$. Thus, to ensure that $\hat{\hat{F}}_K^{ded}$ approximates F well, widely spread deductible choices are required. If K goes to ∞, but the additional included deductible choices are all concentrated in a particular region of the interval $(0, N)$, we cannot improve the approximation of the function F in the region where fewer deductible choices are located.

When implementing the function $\hat{\hat{F}}_K^{ded}$, the values $z_{K,1}, \ldots, z_{K,K}$ are sometimes negative, especially for small K. Because of the negativity of some of the values $z_{K,1}, \ldots, z_{K,K}$, the function $\hat{\hat{F}}_K^{ded}$ is not always non-decreasing, which contradicts the definition of a cumulative distribution function. To overcome this problem, we derive a non-decreasing cumulative distribution function $\hat{F}_K^{ded}$ using $\hat{\hat{F}}_K^{ded}$.[10] First, we run the following algorithm:

Algorithm 1: Determining $\hat{x}_1, \ldots, \hat{x}_{\hat{K}}$

$\hat{x}_0 = \varepsilon < 0$

$i = 1$

$y = 0$

while $y < 1$ **do**

  $\hat{x}_i = \inf\{x \geq \hat{x}_{i-1} \mid \hat{\hat{F}}_k^{ded}(x) > y\}$

  $\hat{x}_{i+1} = \min(\sup\{x \geq \hat{x}_i \mid \hat{\hat{f}}_K^{ded}(z) \geq 0 \text{ and } \hat{\hat{F}}_K^{ded}(z) \leq 1 \text{ for all } z \in [\hat{x}_i, x]\}, N)$

  $y = \hat{\hat{F}}_K^{ded}(\hat{x}_{i+1})$

  $i = i + 2$

end

$\hat{K} = i - 1$

Given $\hat{x}_1, \ldots, \hat{x}_{\hat{K}}$, we define $\hat{F}_K^{ded}$ in the following way:

$$\hat{F}_K^{ded}(x) = \begin{cases} 0 & \text{if } x < \hat{x}_1 \\ \hat{\hat{F}}_K^{ded}(x) & \text{if } \hat{x}_1 \leq x \leq \hat{x}_2 \\ \hat{\hat{F}}_K^{ded}(\hat{x}_2) & \text{if } \hat{x}_2 \leq x \leq \hat{x}_3 \\ \hat{\hat{F}}_K^{ded}(x) & \text{if } \hat{x}_3 \leq x \leq \hat{x}_4 \\ \hat{\hat{F}}_K^{ded}(\hat{x}_4) & \text{if } \hat{x}_4 \leq x \leq \hat{x}_5 \\ & \vdots \\ 1 & \text{if } x \geq \hat{x}_{\hat{K}} \end{cases}.$$

Like the function $\hat{\hat{F}}_K^{ded}$, this distribution function converges to F under the conditions specified in Theorem 3.1.

[10] We have analyzed several procedures to overcome the problem that $\hat{\hat{F}}_k^{ded}$ is not always non-decreasing. We chose the given method because it produces lower expected costs of choosing a non-optimal deductible than other procedures when estimating optimal deductible choices.

From the perspective of our analysis, adding additional deductible policies is always beneficial because doing so improves the approximation $\hat{F}_K^{ded}$, especially when the deductible levels are widely distributed across the loss range. This contradicts psychological research, which shows that too much choice can lead to poorer decisions, as the effort to compare different contracts increases (cf., e.g., Iyengar and Lepper, 2000; Iyengar and Kamenica, 2010). This psychological phenomenon can be overcome if decision aids are used that present policyholders with only the most advantageous contracts instead of all contracts. In this way, distributional information from additional deductible policies can be used without overwhelming the decision-making ability of the individual.

The next question we will address is how to obtain a good approximation of F using the empirical distribution function $\hat{F}_n$ and the function $\hat{F}_K^{ded}$. The potential policyholder forms a weighted average of $\hat{F}_n$ and $\hat{F}_K^{ded}$ given by

$$\hat{F}_{n,K}^{*} = \kappa_{n,K}\hat{F}_K^{ded} + (1-\kappa_{n,K})\hat{F}_n,$$

where $0 \leq \kappa_{n,K} \leq 1$. For the probability distribution $\hat{F}_{n,K}^{*}$, we define a function that measures the deviation from the actual probability distribution F. However, before presenting the measure used in this paper, we need to define the following variables:

$$\begin{aligned} \mu_K^{ded} &= \int_0^N l \; d\hat{F}_K^{ded}(l), \\ \bar{L} &= \frac{1}{n}\sum_{i=1}^{n} L_i, \\ \sigma_{ded,K}^2 &= \int_0^N (l-\mu)^2 \; d\hat{F}_K^{ded}(l), \\ \sigma_n^2 &= \frac{1}{n}\sum_{i=1}^{n}(L_i-\mu)^2. \end{aligned}$$

Given these definitions, the measure we will use is given by

$$\begin{aligned} d_{n,K}\left(\kappa_{n,K}\right) = &\left(\kappa_{n,K}\mu_K^{ded} + (1-\kappa_{n,K})\bar{L} - \mu\right)^2 \\ &+ \beta\left(\kappa_{n,K}\sigma_{ded,K}^2 + (1-\kappa_{n,K})\sigma_n^2 - \sigma^2\right)^2, \end{aligned} \tag{6}$$

where $\beta > 0$. In the first part of the sum shown in (6), the squared deviation of the mean of $\hat{F}_{n,K}^{*}$ from the actual mean μ is calculated. In the second part, we determine the squared deviation of the variance estimate $\kappa_{n,K}\sigma_{ded,K}^2 + (1-\kappa_{n,K})\sigma_n^2$ from the actual variance σ^2 and weight it with the coefficient β, which measures the policyholder's aversion to deviations from the actual variance σ^2.

Next, we compute the value $0 \leq \hat{\kappa}^*_{n,K} \leq 1$ for which $d_{n,K}(\kappa_{n,K})$ is minimized. To do this, we derive $d_{n,K}(\kappa_{n,K})$ by $\kappa_{n,K}$ and set the corresponding derivative to zero. Solving for $\kappa_{n,K}$ then leads to the following solution:

$$\tilde{\kappa}_{n,K} = \frac{(\mu_K^{ded} - \bar{L})(\mu - \bar{L}) + \beta(\sigma^2_{ded,K} - \sigma^2_n)(\sigma^2 - \sigma^2_n)}{(\mu_K^{ded} - \bar{L})^2 + \beta(\sigma^2_{ded,K} - \sigma^2_n)^2}.$$

Given this solution, we obtain the following formula for the weight $\kappa^*_{n,K}$:

$$\hat{\kappa}^*_{n,K} = \begin{cases} 0 & \text{if } \tilde{\kappa}_{n,K} < 0 \\ \tilde{\kappa}_{n,K} & \text{if } 0 \leq \tilde{\kappa}_{n,K} \leq 1 \\ 1 & \text{if } \tilde{\kappa}_{n,K} > 1 \end{cases} . \tag{7}$$

For a better understanding of $\hat{\kappa}^*_{n,K}$, we present selected properties of this weight in the following lemma:

Lemma 3.2. *The weight $\hat{\kappa}^*_{n,K}$ given by (7) has the following properties:*

(i) Let us assume that K is fixed. Then, $\lim_{n\to\infty} \hat{\kappa}^*_{n,K} = 0$ $\mathbb{P}$-almost surely.

(ii) Let $(D_K)_{K\in\mathbb{N}}$ be a sequence of deductible choices and denote by $\bar{D}_K$ the partition of the first K deductible choices of this sequence. Further assume that n is fixed and $\lim_{K\to\infty} \delta(\bar{D}_K) = 0$. Then, $\lim_{K\to\infty} \hat{\kappa}^*_{n,K} = 1$.

(iii) Assume that $\beta = 0$. If $\mu_K^{ded} < \mu$, the cumulative distribution function of $\hat{\kappa}^*_{n,K}$ can be approximated by

$$G_1(x) = \begin{cases} 0 & \textit{if } x < 0 \\ \frac{1}{2} - \Phi\left(\frac{\sqrt{n}(\mu_K^{ded} - \mu)}{\sigma}\right) & \textit{if } x = 0 \\ \Phi\left(\frac{x}{x-1}\frac{\sqrt{n}(\mu_K^{ded} - \mu)}{\sigma}\right) - \Phi\left(\frac{\sqrt{n}(\mu_K^{ded} - \mu)}{\sigma}\right) & \textit{if } 0 < x < 1 \\ 1 & \textit{if } x \geq 1 \end{cases} \tag{8}$$

for large n, where Φ denotes the cumulative distribution function of the standard normal distribution.

If $\mu_K^{ded} > \mu$, the cumulative distribution function of $\hat{\kappa}_{n,K}^*$ may be approximated by

$$G_2(x) = \begin{cases} 0 & \text{if } x < 0 \\ \Phi\left(\frac{\sqrt{n}(\mu_K^{ded}-\mu)}{\sigma}\right) - \frac{1}{2} & \text{if } x = 0 \\ \Phi\left(\frac{\sqrt{n}(\mu_K^{ded}-\mu)}{\sigma}\right) - \Phi\left(\frac{x}{x-1}\frac{\sqrt{n}(\mu_K^{ded}-\mu)}{\sigma}\right) & \text{if } 0 < x < 1 \\ 1 & \text{if } x \geq 1 \end{cases} \tag{9}$$

for large n, where Φ denotes the cumulative distribution function of the standard normal distribution. For $\mu_K^{ded} = \mu$, we get $\hat{\kappa}_{n,K}^* = 1$.

In property (i) in Lemma 3.2, the case in which n goes to ∞ corresponds to a scenario where the policyholder can derive the exact form of F using the empirical distribution function. This is the reason why she should rely only on the empirical distribution function corresponding to the case where $\hat{\kappa}_{n,K}^* = 0$. Property (ii) describes the case in which the policyholder uses only the information derived from available insurance quotes to determine her choice of deductible. This is reasonable since the probability distribution function $\hat{F}_K^{ded}$ converges to F for $K \to \infty$ according to Theorem 3.1.

If we consider a measure $d_{n,K}$ for which $\beta = 0$ holds, it is possible to derive approximate cumulative distribution functions for $\hat{\kappa}_{n,K}^*$, as shown in (iii) in Lemma 3.2. Looking more closely at the formulas in (8) and (9), we note two interesting facts. First, the formulas show that the closer μ_K^{ded} is to μ, the more the probability mass is centered around $\hat{\kappa}_{n,K}^* = 1$. This makes sense because for $\beta = 0$, the measure $d_{n,K}$ only measures the deviation from the mean μ, not the deviation from σ^2. Thus, as μ_k^{ded} approaches μ, the probability mass should shift toward $\hat{\kappa}_{n,K}^*$. Moreover, we see that the probability mass around $\hat{\kappa}_{n,K}^* = 0$ increases as n increases. According to the central limit theorem, this is reasonable because $\bar{L}$ is more concentrated around the true mean μ for larger n.

Having derived an approximation for the loss distribution for n loss observations and K available deductible levels, let us now examine how the policyholder finds an approximation for her optimal deductible level D_K^*. For this purpose, we assume that she uses $\hat{F}_{n,K}^*$ as an approximation for F and knows the exact functional form of the utility function u.[11]

Under the given assumptions, the policyholder's optimal deductible choice for n observations

[11] In Section 4, we will extend the following discussion to the case where we can only approximate u.

and K available deductible levels is given by

$$\hat{D}^*_{n,K} = \underset{D \in \{D_1,\dots,D_K,N\}}{\arg\max} \int_0^N u\left(Y_D\left(l\right)\right)\ d\hat{F}^*_{n,K}(l).$$

This expression converges $\mathbb{P}$-almost surely to the optimal deductible level under full information D^*_K as n goes to ∞. The rate of convergence of the sequence $(D^*_{n,K})_{n\in\mathbb{N}}$ depends on several factors, which we will analyze below. For this analysis, we define

$$U\left(D\right) = \int_0^N u\left(Y_D\left(l\right)\right) dF(l),$$
$$\hat{U}_{n,K}\left(D\right) = \int_0^N u\left(Y_D\left(l\right)\right) d\hat{F}^*_{n,K}(l).$$

By Lemma 3.2, $\lim_{n\to\infty} \hat{\kappa}^*_{n,K} = 0$ $\mathbb{P}$-almost surely. Thus, for large n, $\hat{F}^*_{n,K} = \hat{F}_n$ holds approximately. As a result, it follows that

$$\begin{pmatrix} \hat{U}_{n,K}\left(D_1\right) \\ \vdots \\ \hat{U}_{n,K}\left(D_K\right) \\ \hat{U}_{n,K}\left(N\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n u\left(Y_{D_1}\left(L_i\right)\right) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n u\left(Y_{D_K}\left(L_i\right)\right) \\ \frac{1}{n}\sum_{i=1}^n u\left(Y_N\left(L_i\right)\right) \end{pmatrix}$$

holds approximately for large n. Using the multivariate central limit theorem (cf., e.g., Klenke, 2013), we obtain the following approximation for large n:

$$\begin{pmatrix} \hat{U}_{n,K}\left(D_1\right) \\ \vdots \\ \hat{U}_{n,K}\left(D_K\right) \\ \hat{U}_{n,K}\left(N\right) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} U\left(D_1\right) \\ \vdots \\ U\left(D_K\right) \\ U\left(N\right) \end{pmatrix}, \frac{1}{n}\Sigma\right), \tag{10}$$

where the elements of the covariance Matrix Σ are

$$\Sigma_{i,j} = \begin{cases} \mathrm{Var}(u(Y_{D_i}(L))) & \text{if } i = j \\ \mathrm{Cov}(u(Y_{D_i}(L)), u(Y_{D_j}(L))) & \text{if } i \neq j \end{cases}.$$

Thereby, D_{K+1} is set to N. Using this approximation, we obtain the following theorem:

Theorem 3.3. *Assume that the policyholder knows the exact functional form of her utility function u. Then, for large n, the distribution of the optimal deductible $\hat{D}^*_{n,K}$ is approximately* *given by*

$$\mathbb{P}\left(\hat{D}^*_{n,K} = D\right) = \begin{cases} \Phi\left(\sqrt{n}\frac{U(D_1)-U(D_2)}{\sqrt{\Sigma_{1,1}+\Sigma_{2,2}-2\Sigma_{1,2}}}\right) & \text{if } D = D_1 \\ \Phi_2\left(\sqrt{n}\tilde{\Sigma}_k^{-\frac{1}{2}}\begin{pmatrix} U(D_k)-U(D_{k-1}) \\ U(D_k)-U(D_{k+1})\end{pmatrix}\right) & \text{if } D = D_k \in \{D_2,\ldots,D_K\} \\ \Phi\left(\sqrt{n}\frac{U(N)-U(D_K)}{\sqrt{\Sigma_{K,K}+\Sigma_{K+1,K+1}-2\Sigma_{K,K+1}}}\right) & \text{if } D = N \end{cases},$$

where Φ is the cumulative distribution function of the standard normal distribution, Φ_2 is the cumulative distribution function of the two-dimensional standard normal distribution, and $\tilde{\Sigma}_k = B^{(k)}\Sigma(B^{(k)})^T$ is a $2 \times (K+1)$ matrix $B^{(k)}$ with

$$B^{(k)}_{i,j} = \begin{cases} 0 & \text{if } j \neq \{k-1,k,k+1\} \text{ or } (i,j) \in \{(1,k+1),(2,k-1)\} \\ 1 & \text{if } (i,j) \in \{(1,k-1),(2,k+1)\} \\ -1 & \text{if } j = k \end{cases}.$$

A look at the approximate distribution of $\hat{D}^*_{n,K}$ for large n given in Theorem 3.3 shows that the probabilities $\mathbb{P}(\hat{D}^*_{n,K} = D_1), \ldots, \mathbb{P}(\hat{D}^*_{n,K} = D_K)$ and $\mathbb{P}(\hat{D}^*_{n,K} = N)$ depend on the factors $\sqrt{n}\gamma_{D_1}, \ldots, \sqrt{n}\gamma_{D_K}, \sqrt{n}\gamma_N$, where

$$\gamma_D = \begin{cases} \frac{U(D_1)-U(D_2)}{\sqrt{\Sigma_{1,1}+\Sigma_{2,2}-2\Sigma_{1,2}}} & \text{if } D = D_1 \\ \tilde{\Sigma}_k^{-\frac{1}{2}}\begin{pmatrix} U(D_k)-U(D_{k-1}) \\ U(D_k)-U(D_{k+1})\end{pmatrix} & \text{if } D = D_k \in \{D_2,\ldots,D_K\} \\ \frac{U(N)-U(D_K)}{\sqrt{\Sigma_{1,1}+\Sigma_{K,K}-2\Sigma_{K+1,K+1}}} & \text{if } D = N \end{cases}.$$

Thereby, $\tilde{\Sigma}_k$ is the same as in Theorem 3.3. For the optimal deductible level D^*_K as given in (1), we obtain $\gamma_{D^*_K} \geq 0$. If $D \in \{D_1,\ldots,D_K,N\} \setminus \{D^*_K\}$, $\gamma_D \leq 0$ if $D \in \{D_1, N\}$ and $\gamma_D \notin \mathbb{R}^2_{\geq 0}$ [12] if $D \in \{D_2,\ldots,D_K\}$. Thus, the higher n is, the higher is the probability $\mathbb{P}(\hat{D}^*_{n,K} = D^*_K)$ of choosing the utility-maximizing deductible level D^*_K. Moreover, when the standardized utility differences $\gamma_{D_1},\ldots,\gamma_{D_K},\gamma_N$ are closer to 0 or $(0,0)^T$, the probability $\mathbb{P}(\hat{D}^*_{n,K} = D^*_K)$ for a fixed n decreases.

Moreover, we are able to derive an approximation of the expected costs of choosing a non-optimal deductible level $D^*_{n,K} \in \{D_1,\ldots,D_K,N\} \setminus \{D^*_K\}$ using the distribution derived in Theorem 3.3. Before we determine the expected costs, we need to define a cost measure. For

[12] $\mathbb{R}^2_{\geq 0}$ denotes the set $\{(x,y) \in \mathbb{R}^2 \mid x \geq 0 \text{ and } y \geq 0\}$.

this purpose, we define the cost function $c : \{D_1, \dots, D_K, N\} \to \mathbb{R}$, for which $c(D)$ is defined as the value $c \in \mathbb{R}$ that solves the following equation:

$$\mathbb{E}[u(Y(D_K^*) - c)] = \mathbb{E}[u(Y(D))]. \tag{11}$$

Since D_K^* is the optimal deductible choice, $c(D_K^*) = 0$ holds. Moreover, the concavity of u ensures that $c(D) > 0$ for $D \in \{D_1, \dots, D_K, N\} \setminus \{D_K^*\}$.

Given this cost function, we can approximate the expected costs associated with choosing a non-optimal deductible using the approximate distribution of the random variable $\hat{D}^*_{n,K}$ derived in Theorem 3.3. These expected costs are given by

$$\mathbb{E}[c(\hat{D}^*_{n,K})] = \sum_{i=1}^{K} \mathbb{P}(\hat{D}^*_{n,K} = D_i)c(D_i) + \mathbb{P}(\hat{D}^*_{n,K} = N)c(N). \tag{12}$$

Since $\hat{D}^*_{n,k}$ converges $\mathbb{P}$-almost surely to D_K^* for $n \to \infty$, $\mathbb{E}[c(\hat{D}^*_{n,K})]$ converges to 0. Thus, with full information, which corresponds to n going to ∞, there is no cost associated with incorrectly chosen deductibles. For finite n, however, the costs can be positive. In this case, the expected costs given in (12) can be considered as the benchmark that potential decision rules which seek to approximate the optimal deductible choice have to undercut.

4 Preference Uncertainty under Limited Information

So far, we have ruled out preference uncertainty by assuming that the policyholder knows the exact functional form of her utility function u. In this section, we consider the optimal deductible choice in the case where preference uncertainty exists. That is, the policyholders knows only the functional values $0 = u(W - N) < u(x_1) < \cdots < u(x_m) < u(W) = 1$ of her utility function u.

Using these functional values, the policyholder must derive an approximation of the utility function u. We assume that she approximates u by the piecewise linear function $u_m : [W - N, W] \to [0, 1]$ with $u_m(z) = u_m^{(h)}(\mathbf{x}_m, z)$, where $\mathbf{x}_m = (x_1, \dots, x_m)$ and $u_m^{(h)} : \mathcal{X}^m \times [W -$

$N, W] \to [0,1]$ with

$$u_m^{(h)}(\mathbf{z}_m, z) = \begin{cases} \frac{u(z_1)}{z_1}(z-(W-N)) & \text{if } W-N \leq z \leq z_1 \\ u(z_1) + \frac{u(z_2)-u(z_1)}{z_2-z_1}(z-z_1) & \text{if } z_1 < z \leq z_2 \\ \vdots & \vdots \\ u(z_{m-1}) + \frac{u(z_m)-u(z_{m-1})}{z_m-z_{m-1}}(z-z_{m-1}) & \text{if } z_{m-1} < z \leq z_m \\ u(z_m) + \frac{1-u(z_m)}{W-z_m}(z-z_m) & \text{if } z_m < z \leq W \end{cases},$$

where $\mathcal{X}^m = \{\mathbf{x}_m \in \mathbb{R}^m | W-N < x_1 < \cdots < x_m < W\}$[13]. An illustration of this piecewise linear function is shown in Figure 3. If $m=0$, u_0 is given by $u_0(z) = \frac{1}{N}(z-(W-N))$, which is the utility function of a risk-neutral policyholder.

Unlike u, u_m is not everywhere twice differentiable, because u_m is not differentiable at the points $x_1, \ldots, x_m$. To ensure differentiability, we can approximate u_m by a function $\tilde{u}_m$ given by $\tilde{u}_m : [W-N, W] \to [0,1]$ with $\tilde{u}_m(z) = \tilde{u}_m^{(h)}(\mathbf{x}_m, z)$, where $\tilde{u}_m^{(h)} : \mathcal{X}^m \times [W-N, W] \to [0,1]$ with

$$\tilde{u}_m(\mathbf{z}_m, z) = \begin{cases} u_m^{(h)}(\mathbf{z}_m, z) & \text{if } W-N \leq z \leq z_1 - t \\ h_1(z) & \text{if } z_1 - t < z \leq z_1 + t \\ u_m^{(h)}(\mathbf{z}_m, z) & \text{if } z_1 + t < z \leq z_2 - t \\ \vdots & \vdots \\ u_m^{(h)}(\mathbf{z}_m, z) & \text{if } z_{m-1} + t < z \leq z_m - t \\ h_m(z) & \text{if } z_m - t < z \leq z_m + t \\ u_m^{(h)}(\mathbf{z}_m, z) & \text{if } z_m + t < z \leq W \end{cases},$$

where $t > 0$ is arbitrarily small and the functions $h_1, \ldots, h_m$ are increasing, concave, twice differentiable, and tangent to $f(z) = u_m^{(h)}(\mathbf{z}_m, z)$ at the points $z_1-t, z_1+t, z_2-t, \ldots, z_m-t, z_m+t$, as illustrated in Figure 3 for $\mathbf{z}_m = \mathbf{x}_m$.[14] When $m=0$, we get $u_0 = \tilde{u}_0$.

First, we see that $\tilde{u}_m(x)$ converges to $u(x)$ for $m \to \infty$ for any $x \in [W-N, W]$ as desired. The next question we seek to address is how u and $\tilde{u}_m$ are related for a fixed $m \in \mathbb{N}$. Obviously, for u, the relation $u = g(\tilde{u}_m(x))$ holds with $g(x) = u(\tilde{u}_m^{-1}(x))$. Pratt (1964) has shown that u would be a uniformly more risk-averse utility function than $\tilde{u}_m$ if $g' > 0$ and $g'' < 0$. Unfortunately, the condition $g'' < 0$ does not hold for all $x \in [0,1]$ for $m > 0$, so we cannot say that u is uniformly more risk-averse than $\tilde{u}_m$. For this reason, we cannot derive general results about

[13]redIn the rest of this paper, we will denote the set $\mathcal{X}^1$ by $\mathcal{X}$ to reduce the number of indices used.
[14]The idea for this approximation has been adopted from Zheng (2020).

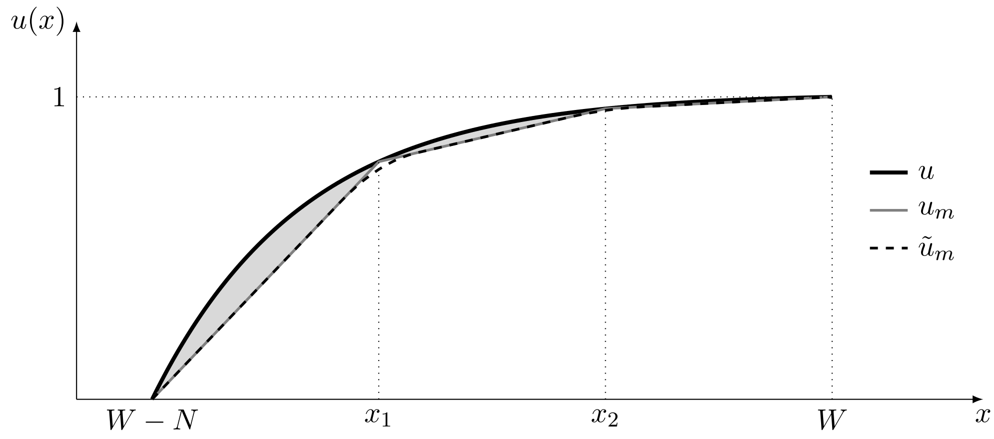


Figure 3: Illustration of the functions u_m and $\tilde{u}_m$ for $\mathbf{x}_2 = (x_1, x_2)$

how the optimal deductible choice under the utility function $\tilde{u}_m$ given by

$$\hat{D}^*_{m,K} = \underset{D \in \{D_1,\ldots,D_K,N\}}{\arg\max} \; \mathbb{E}\left[\tilde{u}_m\left(Y_D\left(L\right)\right)\right]$$

deviates from the optimal deductible choice D^*_K under u. Next, we consider the deviation of the function $\tilde{u}_m$ from u, shown as the gray area in Figure 3. In general, the choice of the values $\mathbf{x}^m \in \mathcal{X}^m$ has a strong impact on this deviation. Thus, the goal of the policyholder is to find the vector $\mathbf{x}^* \in \mathcal{X}^m$ for which

$$\mathbf{x}^* = (x^*_1, \ldots, x^*_m) = \underset{\mathbf{x}_m \in \mathcal{X}^m}{\arg\min} \int_{W-N}^{W} (u(z) - \tilde{u}_m(z))^2 dz. \tag{13}$$

However, the policyholder does not know u, so she cannot solve the minimization problem in (13).

To address the problem that the policyholder is unable to solve (13), we assume that the survey designer (e.g., a consumer protection agency) knows the distribution of utility functions across the population of potential policyholders.[15] Using this knowledge, the survey designer is able to define a stepwise procedure that determines a vector $\mathbf{x}^{**} = (x^{**}_1, \ldots, x^{**}_m) \in \mathcal{X}^m$ that optimally exploits the information that the policyholder reveals. In the following paragraphs, we present this procedure in detail.

First, assume that the set of all possible utility functions is described by the probability space $(\mathcal{U}, \mathcal{F}, \mathbb{P}_{\mathcal{U}})$. Since the policyholder is part of the population, $u \in \mathcal{U}$ obviously holds. An example

[15]Recall that the PE method requires the policyholder to answer a survey in which she is presented with hypothetical decision situations and thereby identifies her indifference probabilities (cf. Figure 1).

of the set $\mathcal{U}$ is

$$\mathcal{U}_Z = \left\{ p(x) = \sum_{i=0}^{n} a_i x^i \middle| n \leq Z, a_0, \dots, a_n \in \mathbb{Q}, p(W-N) = 0, p(W) = 1, p' > 0, p'' \leq 0 \right\},$$

where $Z \in \mathbb{N}$ – that is, the set of all polynomials of degree at most Z that have rational coefficients and satisfy the necessary conditions $p(W-N) = 0$, $p(W) = 1$, and $p' > 0$ and $p'' \leq 0$. For example, Murray (1972) considers polynomials of degree $Z = 4$ which were fitted to survey data.

Knowing the distribution of utility functions in the entire population of potential policyholders, the survey designer can derive the value $\tilde{x}_1^{**} \in \mathcal{X}$ for which

$$\tilde{x}_1^{**} = \arg\min_{x \in \mathcal{X}} \sum_{v \in \mathcal{U}} \mathbb{P}_{\mathcal{U}}(v) \int_{W-N}^{W} (v(z) - \tilde{v}_1(z))^2 dz. \tag{14}$$

That is, the survey designer chooses the value $x \in \mathcal{X}$ such that the average squared deviation of the approximation function from the true utility function is minimized over the entire population of potential policyholders. In the first step, the policyholder determines the functional value $u(\tilde{x}_1^{**})$. Then, the survey designer chooses $\tilde{x}_2^{**} \in \mathcal{X}$ such that

$$\tilde{x}_2^{**} = \arg\min_{x \in \mathcal{X}} \sum_{v \in \mathcal{U}} \mathbb{P}_{\mathcal{U}|u(\tilde{x}_1^{**})}(v) \int_{W-N}^{W} \left(v(z) - \tilde{v}_2^{(h)}(r((\tilde{x}_1^{**}, x)), z)\right)^2 dz,$$

where the function r sorts the corresponding vector in ascending order. In this equation, $\mathbb{P}_{\mathcal{U}|u(\tilde{x}_1^{**})}$ denotes the conditional expectation $\mathbb{P}_{\mathcal{U}}(A|\tilde{U}_1)$, where $A \in \mathcal{F}$ and $\tilde{U}_1 = \{v \in \mathcal{U} | v(\tilde{x}_1^{**}) = u(\tilde{x}_1^{**})\}$. Following this procedure, the value $\tilde{x}_m^{**} \in \mathcal{X}$ is given by

$$\tilde{x}_m^{**} = \arg\min_{x \in \mathcal{X}} \sum_{v \in \mathcal{U}} \mathbb{P}_{\mathcal{U}|\tilde{\mathbf{u}}_{m-1}}(v) \int_{W-N}^{W} \left(v(z) - \tilde{v}_m^{(h)}(r((\tilde{x}_1^{**}, \dots, \tilde{x}_{m-1}^{**}, x)), z)\right)^2 dz, \tag{15}$$

where $\tilde{\mathbf{u}}_{m-1} = (u(\tilde{x}_1^{**}), \dots, u(\tilde{x}_{m-1}^{**}))$. In this equation, $\mathbb{P}_{\mathcal{U}|\tilde{\mathbf{u}}_{m-1}}$ denotes the conditional expectation $\mathbb{P}_{\mathcal{U}}(A|\tilde{U}_{m-1})$, where $A \in \mathcal{F}$ and

$$\tilde{U}_{m-1} = \{v \in \mathcal{U} | v(\tilde{x}_1^{**}) = u(\tilde{x}_1^{**}), \dots, v(\tilde{x}_{m-1}^{**}) = u(\tilde{x}_{m-1}^{**})\}.$$

Finally, the resulting vector is sorted such that the final survey vector for the policyholder is given by $\mathbf{x}^{**} = (x_1^{**}, \dots, x_m^{**}) = r((\tilde{x}_1^{**}, \dots, \tilde{x}_m^{**}))$. In general, $\mathbf{x}^{**}$ and $\mathbf{x}^{*}$ differ in that $\mathbf{x}^{*}$ builds on the knowledge of the exact functional form of u, whereas $\mathbf{x}^{**}$ uses only the partial information offered by the policyholder in revealing certain functional values of u to the survey designer.

In comparison to commonly used survey designs for assessing utility functions (cf., e.g., Murray, 1972; Hershey et al., 1982; Hershey and Schoemaker, 1985; Bleichrodt et al., 2001), our stepwise procedure has two major advantages. First, it determines the survey vector $\mathbf{x} \in \mathcal{X}^m$ iteratively by leveraging available population information while other studies do not account for population information and use a pre-defined survey vector $\mathbf{x} \in \mathcal{X}^m$. Second, by focusing on squared deviations between utility functions, it covers the curvature of a utility function more accurately than other designs.

In the next step, we analyze how the policyholder finds an approximation for her optimal deductible level D_K^*. For this purpose, we assume that she uses $\hat{F}_{n,K}^*$ as an approximation for F and $\tilde{u}_m$ as an approximation for u. Under these assumptions, the policyholder's optimal deductible choice for n loss observations, K available deductible levels $D_1, \ldots, D_K$, and m functional values $u(x_1), \ldots, u(x_m)$ is given by

$$\hat{D}_{m,n,K}^* = \underset{D \in \{D_1, \ldots, D_K, N\}}{\arg\max} \int_0^N \tilde{u}_m \left(Y_D\left(l\right)\right) \; d\hat{F}_{n,K}^*(l). \tag{16}$$

This expression converges $\mathbb{P}$-almost surely to the optimal deductible level under full information D_K^* when n and m go to ∞.

Next, we derive the counterpart of Theorem 3.3 for the case where preference uncertainty exists. For this purpose, we define

$$U_m\left(D\right) = \int_0^N \tilde{u}_m \left(Y_D\left(l\right)\right) dF(l),$$
$$\hat{U}_{m,n,K}\left(D\right) = \int_0^N \tilde{u}_m \left(Y_D\left(l\right)\right) d\hat{F}_{n,K}^*(l).$$

According to Lemma 3.2, $\lim_{n\to\infty} \hat{\kappa}_{n,K}^* = 0$ $\mathbb{P}$-almost surely. Thus, for large n, $\hat{F}_{n,K}^* = \hat{F}_n$ holds approximately. As a result, it follows that

$$\begin{pmatrix} \hat{U}_{m,n,K}\left(D_1\right) \\ \vdots \\ \hat{U}_{m,n,K}\left(D_K\right) \\ \hat{U}_{m,n,K}\left(N\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \tilde{u}_m\left(Y_{D_1}\left(L_i\right)\right) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n \tilde{u}_m\left(Y_{D_K}\left(L_i\right)\right) \\ \frac{1}{n}\sum_{i=1}^n \tilde{u}_m\left(Y_N\left(L_i\right)\right) \end{pmatrix}$$

holds approximately for large n. Using the multivariate central limit theorem (cf., e.g., Klenke, 2013), we obtain the following approximation for large n:

$$\begin{pmatrix} \hat{U}_{m,n,K}(D_1) \\ \vdots \\ \hat{U}_{m,n,K}(D_K) \\ \hat{U}_{m,n,K}(N) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} U_m(D_1) \\ \vdots \\ U_m(D_K) \\ U_m(N) \end{pmatrix}, \frac{1}{n}\Sigma^{(m)} \right), \tag{17}$$

where the elements of the covariance Matrix $\Sigma^{(m)}$ are

$$\Sigma^{(m)}_{i,j} = \begin{cases} \mathrm{Var}(\tilde{u}_m(Y_{D_i}(L))) & \text{if } i = j \\ \mathrm{Cov}(\tilde{u}_m(Y_{D_i}(L)), \tilde{u}_m(Y_{D_j}(L))) & \text{if } i \neq j \end{cases}.$$

Thereby, D_{K+1} is set to N. Using this approximation, we can derive the following theorem:

Theorem 4.1. *Suppose the policyholder does not know the exact functional form of her utility function u and observes only m functional values $u(x_1), \ldots, u(x_m)$. Then, for large n, the distribution of the optimal deductible $\hat{D}^*_{m,n,K}$ is approximately given by*

$$\mathbb{P}\left(\hat{D}^*_{m,n,K} = D\right) = \begin{cases} \Phi\left(\sqrt{n}\frac{U_m(D_1)-U_m(D_2)}{\sqrt{\Sigma^{(m)}_{1,1}+\Sigma^{(m)}_{2,2}-2\Sigma^{(m)}_{1,2}}}\right) & \textit{if } D = D_1 \\ \Phi_2\left(\sqrt{n}(\tilde{\Sigma}^{(m)}_k)^{-\frac{1}{2}} \begin{pmatrix} U_m(D_k) - U_m(D_{k-1}) \\ U_m(D_k) - U_m(D_{k+1}) \end{pmatrix}\right) & \textit{if } D = D_k \in \boldsymbol{D}^{-1}_K \\ \Phi\left(\sqrt{n}\frac{U_m(D_K)-U_m(D_{K-1})}{\sqrt{\Sigma^{(m)}_{K,K}+\Sigma^{(m)}_{K+1,K+1}-2\Sigma^{(m)}_{K,K+1}}}\right) & \textit{if } D = N \end{cases},$$

where $\boldsymbol{D}^{-1}_K = \{D_2, \ldots, D_K\}$, Φ denotes the cumulative distribution function of the standard normal distribution, Φ_2 denotes the cumulative distribution function of the two-dimensional standard normal distribution, and $\tilde{\Sigma}^{(m)}_k = B^{(k)}\Sigma^{(m)}(B^{(k)})^T$ for a $2 \times (K+1)$ matrix $B^{(k)}$ with

$$B^{(k)}_{i,j} = \begin{cases} 0 & \textit{if } j \neq \{k-1, k, k+1\} \textit{ or } (i,j) \in \{(1,k+1),(2,k-1)\} \\ 1 & \textit{if } (i,j) \in \{(1,k-1),(2,k+1)\} \\ -1 & \textit{if } j = k \end{cases}.$$

Compared to Theorem 3.3, the probability of choosing the utility maximizing deductible D^*_K – that is, $\mathbb{P}(\hat{D}^*_{m,n,K} = D^*_K)$ does not necessarily converge to 1 as n goes to ∞. This is ensured only when m and n go to ∞. When n is large and stays fixed, the approximate distributions of $\hat{D}^*_{m,n,K}$ and $\hat{D}^*_{n,K}$ are identical when m goes to ∞.

Moreover, it is also possible to derive a version of Theorem 4.1 for the case where the elicitation procedure is biased. Here we obtain an approximation function $\tilde{u}_m^{(b)}$ based on biased functional values $u^{(b)}(x_1), \ldots, u^{(b)}(x_m)$. When the optimal deductible choice under $u^{(b)}$, given by $D_K^{(b)}$ in (2), differs from D_K^*, the probability of choosing the utility-maximizing deductible does not converge to one when m and n go to ∞. Moreover, the distribution of $\hat{D}^*_{m,n,K}$ will be biased toward the non-optimal deductible choice $D_K^{(b)}$. To avoid such deviations and their associated costs, accounting for possible biases in the elicitation method is crucial.

While the n loss observations are exogenously given, the policyholder can choose how many functional values m of her utility function u to obtain using the survey provided by the survey designer. The sequence of functional values $(u(x_m^{**}))_{m\in\mathbb{N}}$ that she obtains is determined by the scheme presented in (14) and (15). We assume that obtaining these values involves a cost modeled by a function $\bar{c} : \bar{\mathcal{X}} \to \mathbb{R}$, where $\bar{\mathcal{X}} := \bigcup_{m\in\mathbb{N}} \mathcal{X}^m$. An example of such a cost function is the linear cost function $\bar{c}(x_1, \ldots, x_m) = \bar{c}_0 + \bar{c}_1 m$, which depends only on the sample size m and not on the specific vector $(x_1, \ldots, x_m)$.

While the function $\bar{c}$ covers the cost of obtaining the functional values $u(x_1), \ldots, u(x_m)$, using the information about these values could also lead to a reduction in the cost of choosing a non-optimal deductible level. If the policyholder chooses $m = 0$, her cost is given by

$$\mathbb{E}[c(\hat{D}^*_{0,n,K})] = \sum_{i=1}^{K} \mathbb{P}(\hat{D}^*_{0,n,k} = D_i)c(D_i) + \mathbb{P}(\hat{D}^*_{0,n,K} = N)c(N).$$

If the policyholder chooses $m > 0$, her cost of choosing a non-optimal deductible level is reduced by

$$\bar{G}_{n,K}(m) = \mathbb{E}[c(D^*_{0,n,K})] - \mathbb{E}[c(D^*_{m,n,K})].$$

Taking into account the cost $\bar{c}(x_1^{**}, \ldots, x_m^{**})$, we are able to calculate the value of additional information – an important measure in decision analysis (cf., e.g., LaValle, 1968; Merkhofer, 1977; Keeney, 1982) – in the following way:

$$G_{n,K}(m) = \begin{cases} 0 & \text{if } m = 0 \\ \bar{G}_{n,K}(m) - \bar{c}(x_1^{**}, \ldots, x_m^{**}) & \text{if } m > 0 \end{cases}.$$

Given this formula for the value of additional information, the number of functional values m that maximizes $G_{n,K}$ is given by

$$m^*_{n,K} = \arg\max_{m\in\mathbb{N}_0} G_{n,K}(m).$$

The cost of choosing non-optimal deductibles is then given by $\mathbb{E}[c(\hat{D}_{m^*_{n,k},n,K})]$. This cost can be used as a benchmark for assessing the quality of possible decision rules for the optimal deductible choice.

In addition, it is also possible to derive conditional expectations of the values of additional information across the population of policyholders in the different steps of the survey scheme introduced in (14) and (15). These can be used to define a stopping criterion for the survey scheme by stopping to ask for further functional values if the conditional expectation of the marginal value of additional information is less than or equal to zero.

5 Decision Rules for Supporting Decision Making

After discussing the effects of limited information on the optimal deductible choice, we will examine decision rules for choosing the optimal deductible level under limited information. First, let us consider what properties a good decision rule should have. To this end, we define the information vector

$$I = (\mathbf{D}_K, \mathbf{R}_K) \in \mathbb{R}^{2K}_{>0},$$

where $\mathbf{D}_K = (D_1, \ldots, D_K)$ and $\mathbf{R}_K = (R(D_1), \ldots, R(D_K))$. This information vector covers the information that is independent of the loss realizations $\mathbf{L}_n = (L_1, \ldots, L_n)$ and the utility values $\mathbf{U}_m = (u(x_1), \ldots, u(x_m))$, where the values $W - N < x_1 < \cdots < x_m < N$ are determined by a procedure as presented in (14) and (15). Then we define a decision rule under the information vector I, which we denote by $\hat{\theta}^{(I)}_{m,n}$, in the following way:

Definition 5.1. *A decision rule $\hat{\theta}^{(I)}_{m,n}$ under the information vector I is an estimator of the optimal deductible level D^*_K in (1). It is called*

(i) consistent if $\lim_{m,n\to\infty} \hat{\theta}^{(I)}_{m,n} = D^*_K$ $\mathbb{P}$-almost surely.

(ii) unbiased if $\lim_{m\to\infty} \mathbb{E}[\hat{\theta}^{(I)}_{m,n}] = D^*_K$ holds for all $n \in \mathbb{N}$.

(iii) preference-independent if $\hat{\theta}^{(I)}_{0,n} = \hat{\theta}^{(I)}_{m,n}$ $\mathbb{P}$-almost surely for all $m \in \mathbb{N}$ for a fixed but arbitrary $n \in \mathbb{N}$. If the decision rule is not preference-independent, we call it preference-dependent.

If $u'' < 0$ (i.e., the policyholder is risk-averse) and $D^*_K \neq N$, the decision rule $\hat{\theta}^{(I)}_{m,n}$ is called

(iv) preference-consistent if

$$\lim_{n\to\infty} \mathbb{E}[c(\hat{\theta}^{(I)}_{0,n})] - \lim_{m,n\to\infty} \mathbb{E}[c(\hat{\theta}^{(I)}_{m,n})] > 0, \tag{18}$$

where c is the cost function $c : \{D_1, \dots, D_K, N\} \to \mathbb{R}$ defined in (11). The difference in (18) is called the value of preference information.

Property (i) in Definition 5.1 establishes a link between the settings under full and limited information. If m and n go to ∞, it is possible to derive the exact forms of the loss distribution L and the utility function u. That is, we are in an environment with full information. In this case, a decision rule $\hat{\theta}^{(I)}_{m,n}$ should provide the optimal deductible level D^*_K, as expressed in (i). Moreover, it is desirable that a decision rule leads on average to the optimal deductible level for all numbers of loss realizations $n \in \mathbb{N}$ when there is no preference uncertainty. This is described by property (ii) in Definition 5.1.

Preference-independent decision rules are decision rules that exploit only the information about the loss distribution L. Therefore, they remain the same when $n \in \mathbb{N}$ is fixed and only $m \in \mathbb{N}$ is changed. Preference consistency describes the desirable property that the average cost of incorrectly chosen deductible levels decreases as the information about a policyholder's preferences improves. The first part of the difference in (18) describes the average cost when there is full information about the loss distribution L and no information about the utility function u. The second part covers the expected costs under full information. More information about a policyholder's preferences should lead to a reduction in the average cost of choosing the deductible of a given decision rule. For this reason, the difference in (18) should be positive for an appropriate decision rule. Moreover, we immediately see that preference-independent decision rules cannot be preference-consistent because the difference in (18) is zero for this class of decision rules.

From the policyholder's point of view, a decision rule $\hat{\theta}^{(I)}_{m,n}$ should be consistent and unbiased as well as preference-consistent. In many cases, we can draw on the law of large numbers or the central limit theorem to analyze whether a possible decision rule is (preference-)consistent. However, it is more complicated to determine whether a decision rule is unbiased. In general, we cannot derive closed-form solutions of the probabilities $\mathbb{P}(\hat{\theta}^{(I)}_{m,n} = D_1), \dots, \mathbb{P}(\hat{\theta}^{(I)}_{m,n} = D_K)$ and $\mathbb{P}(\hat{\theta}^{(I)}_{m,n} = N)$ such that it is impossible to obtain a closed-form solution for the expected value

$$\mathbb{E}[\hat{\theta}^{(I)}_{m,n}] = \sum_{k=1}^{K} \mathbb{P}\left(\hat{\theta}^{(I)}_{m,n} = D_k\right) D_k + \mathbb{P}\left(\hat{\theta}^{(I)}_{m,n} = N\right) N. \tag{19}$$

For example, considering the decision rule $\hat{D}^*_{m,n,K}$ in (16), it is not possible to derive its exact distribution for each $n \in \mathbb{N}$, so we cannot analytically determine the expectation in (19). Only for large n, $\mathbb{E}[\hat{D}^*_{m,n,K}]$ can be approximated using the results in Theorem 4.1.

Next, we will present two preference-dependent decision rules on which we will focus in the remainder of our paper. The discussion of their properties according to Definition 5.1 can be found in Appendix B.[16]

Laplace Decision Rule

The Laplace decision rule exploits the preference information $\mathbf{U}_m$ and is defined as follows (cf., e.g., Marchau et al., 2019; Luce and Raiffa, 1989):

$$\hat{\theta}_{m,n}^{(I),1} = \underset{D \in \{D_1,\ldots,D_K,N\}}{\arg\max} \frac{1}{n} \sum_{i=1}^{n} \tilde{u}_m(Y_D(L_i)).$$

Optimal Deductible Decision Rule

The second decision rule, on the other hand, is defined as

$$\hat{\theta}_{m,n}^{(I),2} = \underset{D \in \{D_1,\ldots,D_K,N\}}{\arg\max} \int_0^N \tilde{u}_m(Y_D(l)) d\hat{F}_{n,K}^*(l),$$

which coincides with the random variable $\hat{D}_{m,n,K}^*$ in (16).

From the perspective of Definition 5.1, the two decision rules are of equal quality. However, they differ significantly in terms of costs for small $n \in \mathbb{N}$ because the distribution $F_{n,K}^*$ is much more accurate than $\hat{F}_n$ for small $n \in \mathbb{N}$. Unfortunately, it is not possible to prove this as a general result. For this reason, we will emphasize this aspect in the following numerical illustration.

6 Numerical Illustration of Rule-Based Decision Making

In the last three sections, we have discussed various sources of limited information that a policyholder faces when choosing her optimal deductible level. In what follows, we aim to deepen this discussion by presenting a numerical illustration motivated by the homeowners insurance market in the United States.[17]

For the loss distribution, we assume that the claims frequency is modeled by a Poisson distributed random variable $\bar{N} \sim \text{Poi}(\lambda_P)$. The individual claims amounts are given by independent and identically distributed random variables $C_1, \ldots, C_{\bar{N}} \sim \text{Exp}(\lambda_E)$. Then the cumulative

[16] We also considered different preference-independent decision rules. Results for these decision rules are not presented in this paper but can be obtained by request.

[17] Strictly speaking, our model framework does not perfectly match real homeowners insurance markets. We assume a deductible level per year, whereas in typical homeowners insurance policies the deductible level is set per claim. However, if there is at most one claim per year, the two specifications are identical. This is roughly true for our numerical illustration, since the probability of more than one claim occurring is less than 0.2% (cf. Insurance Information Institute, 2021). Therefore, we believe that our model framework is appropriate and that no adjustments need to be made.

distribution function of the corresponding loss distribution $\tilde{L}$ is

$$\mathbb{P}(\tilde{L} \leq x) = e^{-\lambda_P} + \sum_{k=1}^{\infty} e^{-\lambda_P} \frac{\lambda_P^k}{k!} \Gamma(x; k, \lambda_E),$$

where $\Gamma(x; k, \lambda_E)$ is the cumulative distribution function of a gamma distribution with parameters k and λ_E (cf., e.g., Straub, 1988). As stated in Section 2, we assume that the maximum loss satisfies the condition $0 < N < W$. Therefore, we truncate the random variable $\tilde{L}$ at N so that the distribution of the loss random variable L is as follows:

$$F(x) = \mathbb{P}(L \leq x) = \begin{cases} \frac{\mathbb{P}(\tilde{L} \leq x)}{\mathbb{P}(\tilde{L} \leq N)} & \text{if } x \leq N \\ 1 & \text{else} \end{cases}.$$

For the parameter values λ_P and λ_E, we choose numbers that are close to those observed for the U.S. homeowners insurance market between 2014 and 2018 (see Table 1). In addition, we consider available deductible levels similar to those we observe in this insurance market segment. In terms of assets, we further assume that the policyholder has wealth equal to the per capita disposable income in the United States in 2019.[18] For the loss limit, we assume a value that is about \$10,000 less than the assumed wealth of the policyholder. Finally, for the loading factor λ and the decision weight β, we choose values of 5% and 10%, respectively.

The policyholder's preferences are modeled by a monotonic transformation of a utility function with constant relative risk aversion (CRRA)

$$u_\eta(z) = \begin{cases} ln(z) & \text{if } \eta = 1 \\ \frac{z^{1-\eta}-1}{1-\eta} & \text{else} \end{cases},$$

where $\eta > 0$. To get a sense of the extent of absolute and relative risk aversion in the population, we provide an overview of estimates for these two measures of risk aversion in Table 2.

At first glance, it is apparent that the mean estimates vary considerably across contexts. For example, one recognizes that the average risk preferences derived from gaming contexts appear to be much lower than the average risk preferences measured in insurance contexts. Moreover, even within contexts, there are significant differences in observed risk aversions. In the academic literature, authors discuss several reasons for these substantial differences. First, differences may arise from the selected group of individuals for whom risk preferences are measured (Cohen and Einav, 2007). For example, in Gertner (1993) and Metrick (1995), the individuals analyzed are

[18]Please note that we use per capita disposable income instead of a typical wealth measure such as median wealth per adult, as this is a widely used approach in the academic literature (cf., e.g., Cohen and Einav, 2007).

Parameter	Notation	Value
Parameter for claims frequency	λ_P	5.64%
Parameter for claims amount	λ_E	$\frac{1}{13{,}814}$
Loss limit	N	\$40,000
Available deductible choices	$D_1, \ldots, D_K$	\$500, \$1,000, ..., \$2,500
Loading factor	λ	5%
Weighting factor for the approximation of the loss distribution	β	10%
Wealth	W	\$49,806
Relative risk aversion parameters	$\eta_1, \ldots, \eta_n$	$0.1, 0.5, 1, 2, 3, 5$
Probability weighting factor for gains	γ^+	0.61
Probability weighting factor for losses	γ^-	0.69
Loss aversion parameter	λ_L	2.25

Note: This table contains the parameter values assumed for our numerical illustration. For claims frequency, we choose λ_P such that the expected number of claims matches the average claims frequency in the homeowners insurance market in the United States between 2014 and 2018. According to the Insurance Information Institute (III), the latter was 5.64%. For the claims amount, we choose λ_E such that the expected claims amount of the untruncated probability distribution equals the average claims severity in the homeowners insurance market in the United States between 2014 and 2018. According to the Insurance Information Institute (III), the latter was \$13,814. For the loss limit, we choose a value of \$40,000, which is about \$10,000 less than the policyholder's wealth that we assume for our numerical illustrations. The available deductible levels are similar to those observed in the homeowners insurance market in the United States. The parameter values of the loading factor λ and the weighting factor β have been set to 5% and 10%, respectively. The wealth parameter corresponds to the per capita disposable income in the United States in 2019, which was \$49,806 (cf. Bureau of Economic Analysis (BEA) and the United States Census Bureau (USCB)). For the relative risk aversion parameters $\eta_1, \ldots, \eta_n$, we choose numbers such that each available deductible choice is optimal for at least one type of policyholder. For the parameters γ^+, γ^-, and λ_L, we assume that $\gamma^+ = 0.61$, $\gamma^- = 0.69$, and $\lambda_L = 2.25$, which correspond to the estimations of Tversky and Kahneman (1992).

Table 1: Parameter values of the numerical illustration

television game show participants, who may have different risk preferences than the average population. Second, the differences could be caused by different contexts and stakes influencing the decision behavior (Rabin, 2000; Cohen and Einav, 2007). Third, information frictions could bias measures for risk preferences. Handel and Kolstad (2015) provide empirical evidence for this explanation by showing that the average level of risk aversion is much lower when controlling for information frictions. As shown in Table 2, the average estimates for the baseline model of Handel and Kolstad (2015) that does not control for information frictions are about 20 times higher than the average estimates in the full model that does control for information frictions. Apart from the differences in average risk preferences across different research studies, there is also evidence of large and skewed heterogeneity in risk attitudes (Cohen and Einav, 2007; Ericson and Starc, 2012). This is reflected in the quantiles of the risk aversion parameters presented in Table 2. To account for this heterogeneity and to choose risk attitudes that approximate those presented in Table 2, we construct the probability space of possible utility functions $(\mathcal{U}, \mathcal{F}, \mathbb{P}_{\mathcal{U}})$ in the following way:

$$\mathcal{U} = \left\{ u \middle| u' > 0, u'' < 0, -\frac{cu''(c)}{u'(c)} = \eta \text{ for } \eta \in \{\eta_1, \dots, \eta_n\}, u(W-N) = 0, u(W) = 1 \right\}$$

with $\eta_n > \dots > \eta_1 > 0$, $\mathcal{F} = \mathcal{P}(U)$ and $\mathbb{P}_{\mathcal{U}}(A) = \sum_{i=1}^n p_i \delta_{\eta_i}(A)$, where $0 \le p_i \le 1$, $\sum_{i=1}^n p_i = 1$ and

$$\delta_{\eta_i}(A) = \begin{cases} 1 & \text{if } \eta_i \in A \\ 0 & \text{else} \end{cases}.$$

The relative risk aversion parameters $\eta_1, \dots, \eta_n$ that we use for our numerical illustration are presented in Table 1. For these, we determine the values $p_1, \dots, p_n$ using a cumulative distribution function F_{aCE} with $F_{aCE}(0) = 0$ that has the same quantiles $q_{0.25}, q_{0.5}, \dots, q_{0.9}, q_{0.95}$ as the friction-adjusted benchmark model of Cohen and Einav (2007) and is piecewise linear between 0 and the different quantiles.[19] Given F_{aCE}, $p_1, \dots, p_n$ are determined as follows:

$$\begin{aligned} p_1 &= F_{aCE}\left(\frac{\eta_1 + \eta_2}{2}\right), \\ p_i &= F_{aCE}\left(\frac{\eta_i + \eta_{i+1}}{2}\right) - F_{aCE}\left(\frac{\eta_{i-1} + \eta_i}{2}\right) \text{ for } i \in \{2, \dots, n-1\}, \\ p_n &= 1 - F_{aCE}\left(\frac{\eta_{n-1} + \eta_n}{2}\right). \end{aligned}$$

For the relative risk aversion parameters used in our numerical illustration, we then obtain

[19] For $x \ge q_{0.95}$, we define $F_{aCE}(x) = 0.95 + \frac{0.05}{q_{0.95} - q_{0.9}}(x - q_{0.95})$ if $x \le q_1$, where $q_1 = q_{0.9} + 2(q_{0.95} - q_{0.9})$. For $x \ge q_1$, $F_{aCE}(x) = 1$.

the probabilities $\mathbb{P}_{\mathcal{U}}(\eta = 0.1) = 0.7701$, $\mathbb{P}_{\mathcal{U}}(\eta = 0.5) = 0.0568$, $\mathbb{P}_{\mathcal{U}}(\eta = 1) = 0.0755$, $\mathbb{P}_{\mathcal{U}}(\eta = 2) = 0.0138$, $\mathbb{P}_{\mathcal{U}}(\eta = 3) = 0.0207$ and $\mathbb{P}_{\mathcal{U}}(\eta = 5) = 0.0631$. Using these probabilities, it is possible to calculate the average population costs of the different decision rules. If we had chosen the quantiles of the benchmark model of Cohen and Einav (2007), the procedure used above would lead to the following probabilities: $\mathbb{P}_{\mathcal{U}}^{CE}(\eta = 0.1) = 0.4485$, $\mathbb{P}_{\mathcal{U}}^{CE}(\eta = 0.5) = 0.0758$, $\mathbb{P}_{\mathcal{U}}^{CE}(\eta = 1) = 0.0481$, $\mathbb{P}_{\mathcal{U}}^{CE}(\eta = 2) = 0.0641$, $\mathbb{P}_{\mathcal{U}}^{CE}(\eta = 3) = 0.0962$ and $\mathbb{P}_{\mathcal{U}}^{CE}(\eta = 5) = 0.2673$. That is, the population distribution modeled by $\mathbb{P}_{\mathcal{U}}$ corresponds to a less risk-averse population than the probability distribution given by $\mathbb{P}_{\mathcal{U}}^{CE}$. Whereas we use the probability measure $\mathbb{P}_{\mathcal{U}}$ to determine the survey vector $\mathbf{x}^{**} = (x_1^{**}, \dots, x_m^{**}) \in \mathcal{X}^m$ by using the procedure presented in (14) and (15), we calculate the average population costs for both $\mathbb{P}_{\mathcal{U}}$ and $\mathbb{P}_{\mathcal{U}}^{CE}$.

6.1 Analysis of the Approximation of the Loss Distribution

Having presented the setup of our numerical analysis, let us analyze the probability weight $\hat{\kappa}_{n,K}^*$ in more detail. In Section 3, we only derived results regarding the convergence behavior of $\hat{\kappa}_{n,K}^*$ and the case where $n \in \mathbb{N}$ is large. While these results provide interesting insights from a theoretical point of view, they are of limited value for actual deductible choices, since n is typically low and there are few deductible levels available. Using our numerical illustration, we aim to address a number of questions that remained unclear in Section 3, including: Ceteris paribus, how fast does $\hat{\kappa}_{n,K}^*$ decrease as n increases? Ceteris paribus, how fast does $\hat{\kappa}_{n,K}^*$ increase when $\mathbf{D}_K$ covers a larger range of the interval $(0, N)$?

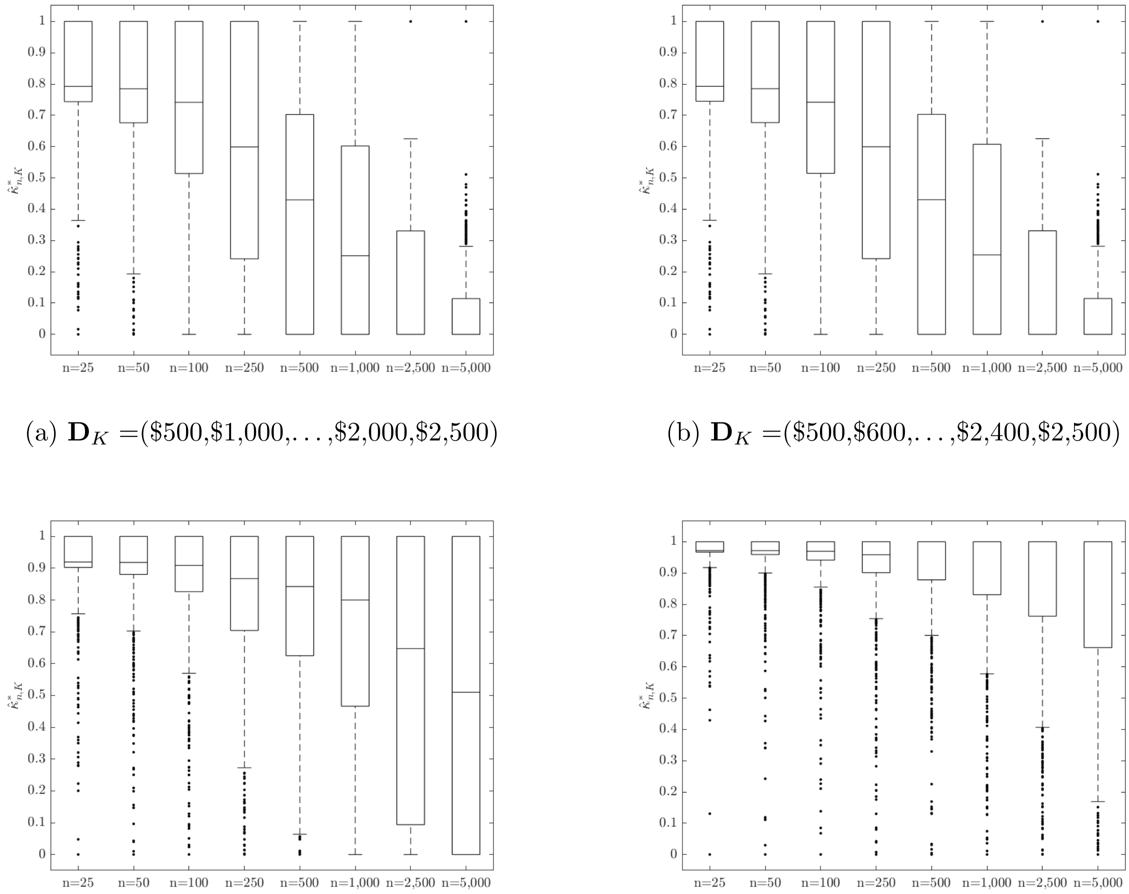
In Figure 4, boxplots illustrate how $\hat{\kappa}_{n,K}^*$ changes as n increases and when $\mathbf{D}_K$ covers a larger range of the interval $(0, N)$. When the deductible levels do not cover a large range (cf. Figure 4 (a)), the median of $\hat{\kappa}_{n,K}^*$ for $n = 25$ is about 0.8 and decreases steadily as n increases. However, even for larger values of n such as $n = 1{,}000$, the median is still significantly different from zero. This means that even with a large number of observations, we use some information from the available deductible policies. One possible reason for this is that we need relatively many observations to fit the tails of the loss distribution well, since the claim frequency is relatively low. Since the cumulative distribution function $\hat{F}_K^{ded}$ covers the tails to some extent, its median probability weight is significantly different from zero even for higher n. If we increase the number of deductibles without changing the range of deductibles, $\hat{\kappa}_{n,K}^*$ does not change substantially (cf. Figure 4 (b)).

From an economic point of view, these are very interesting results. We see that the derived distribution $\hat{F}_K^{ded}$ provides valuable information for approximating the loss distribution even for a relatively narrow deductible range, especially for low n. For example, in the homeowners

Specification	Absolute risk aversion	Relative risk aversion
Mean estimates from game contexts:		
Gertner (1993)	$3.1 \cdot 10^{-4}$	4.44
Metrick (1995)	$6.6 \cdot 10^{-5}$	1.15
Mean estimates from insurance contexts:		
Sydnor (2006)	$1.9 \cdot 10^{-3}$	54.07
Cohen and Einav (2007)	$6.7 \cdot 10^{-3}$	97.22
Handel and Kolstad (2015): Base Model	$1.6 \cdot 10^{-3}$	60.97
Handel and Kolstad (2015): Full Model	$8.6 \cdot 10^{-5}$	3.28
Benchmark model of Cohen and Einav (2007):		
25th percentile	$2.3 \cdot 10^{-6}$	0.03
Median individual	$2.6 \cdot 10^{-5}$	0.37
75th percentile	$2.9 \cdot 10^{-4}$	4.27
90th percentile	$2.7 \cdot 10^{-3}$	39.02
95th percentile	$9.9 \cdot 10^{-3}$	143.27
Friction-adjusted benchmark model of Cohen and Einav (2007):		
25th percentile	$3.0 \cdot 10^{-8}$	0.0011
Median individual	$3.3 \cdot 10^{-7}$	0.0125
75th percentile	$3.7 \cdot 10^{-6}$	0.1410
90th percentile	$3.5 \cdot 10^{-5}$	1.33
95th percentile	$1.3 \cdot 10^{-4}$	4.95

Note: This table shows the estimation results for the coefficients of absolute and relative risk aversion from different research papers. The coefficient of absolute risk aversion (ARA) is converted to US$$^{-1}$. Like Cohen and Einav (2007), we calculate the coefficients of relative risk aversion (RRA) by multiplying the ARA estimates by average annual income. The mean and quantile estimates for Cohen and Einav (2007) are quoted directly from their paper. For the other estimates, we draw on average per capita disposable income (ADI) in the United States as the average income measure, which is consistent with the approach of Cohen and Einav (2007). For the other papers, we use the following ADI figures: ADI in 1987 ($14,330) for Gertner (1993), ADI in 1990 ($17,360) for Metrick (1995), ADI in 2002 ($28,184) for Sydnor (2006), and ADI in 2011 ($38,108) for Handel and Kolstad (2015).
For Handel and Kolstad (2015), we show the results of their baseline model and their full model. The baseline model follows an approach similar to that of Cohen and Einav (2007), while their full model accounts for the information frictions associated with choosing an insurance plan. In addition, we adjust the quantile estimates of the benchmark model of Cohen and Einav (2007) to account for the information frictions. To this end, we use the following approach: First, we divide the mean ARA estimate of the full model of Handel and Kolstad (2015) by the mean ARA estimate of Cohen and Einav (2007) – that is, $\frac{8.6 \cdot 10^{-5}}{6.7 \cdot 10^{-3}}$. We then multiply the quantile ARA estimates in the benchmark model of Cohen and Einav (2007) by this number to obtain the ARA estimates in the section "Friction adjusted benchmark model of Cohen and Einav (2007)". In a final step, we multiply the resulting ARA estimates by the ADI in 2011 ($38,108) to obtain the RRA estimates. This approach serves only as a rough estimate for the distribution of the friction-adjusted RRA estimates. Although this approach is associated with some problems, we nevertheless believe that it provides a good understanding of how large the impact of information frictions on risk estimates can be, especially in the complex decision environment we found in insurance markets.

Table 2: Risk aversion estimates



(a) $\mathbf{D}_K$ =($500,$1,000,. . . ,$2,000,$2,500)

(b) $\mathbf{D}_K$ =($500,$600,. . . ,$2,400,$2,500)

(c) $\mathbf{D}_K$ =($500,$1,000,. . . ,$7,000,$7,500)

(d) $\mathbf{D}_K$ =($500,$1,000,. . . ,$9,500,$10,000)

Note: This figure shows boxplots for the simulated values of the probability weight $\hat{\kappa}^*_{n,K}$ for $\beta = 0.1$ for different $n \in \mathbb{N}$ and different sets of available deductible policies. These results are based on simulations with 1,000 simulation runs.

Figure 4: Distribution of $\hat{\kappa}^*_{n,K}$ for different n and $\mathbf{D}_K$

insurance market, deductibles of $500,. . . ,$2,500 are quite common and could be very helpful in estimating the loss distribution according to our numerical illustration.

When the range of $\mathbf{D}_K$ is expanded, $\hat{\kappa}^*_{n,K}$ increases even further, as Figure 4 (c) and Figure 4 (d) show. For a relatively large range of $\mathbf{D}_K$ (cf. Figure 4 (d)), $\hat{\kappa}^*_{n,K}$ becomes almost unity for low n. Moreover, the median of $\hat{\kappa}^*_{n,K}$ remains close to one even for high n.

6.2 Assessment of Decision Rules for Unbiased Elicitation Processes

We now discuss how well the decision rules in Section 5 estimate the optimal deductible choice D^*_K. To this end, we compare the expected costs of choosing an inappropriate deductible introduced in (12). Table 3 shows the costs associated with such an inappropriate deductible choice for the relative risk aversion parameters used in our numerical analysis. A look at the various figures reveals that the optimal deductible level decreases as relative risk aversion increases, as predicted by theory. Moreover, the closer a given deductible level is to the optimal deductible level D^*_K, the lower is its cost. With the exception of $\eta = 0.1$, the highest cost occurs when no insurance is purchased – that is, when $D =$ \$40,000 is chosen. In this case, the cost can be as high as \$5,639.40.

	\$500	\$1,000	\$1,500	\$2,000	\$2,500	\$40,000
$\eta = 0.1$	14.06	12.88	11.76	10.70	9.71	0*
$\eta = 0.5$	3.27	2.23	1.34	0.60	0*	56.40
$\eta = 1$	1.90	1.04	0.44	0.10	0*	161.73
$\eta = 2$	0.50	0.01	0*	0.48	1.42	499.16
$\eta = 3$	0.13	0*	0.60	1.92	3.95	1,177.60
$\eta = 5$	0*	0.63	2.49	5.59	9.92	5,639.40

Note: This table shows the cost of choosing an inappropriate deductible (measured in US\$) for the relative risk aversion parameters analyzed in our numerical illustration. These costs were calculated using the cost function in (11). When the cost is zero, the chosen deductible is the optimal deductible, which we denote by *.

Table 3: Costs of choosing an inappropriate deductible

In the next step, we can compare the decision rules presented in Section 5 in our numerical analysis framework. To assess the effect of preference uncertainty, we will also provide results for the following estimators:

$$\hat{\theta}_n^{(I),1} = \underset{D\in\{D_1,\dots,D_K,N\}}{\arg\max} \frac{1}{n}\sum_{i=1}^{n} u(Y_D(L_i)),$$
$$\hat{\theta}_n^{(I),2} = \underset{D\in\{D_1,\dots,D_K,N\}}{\arg\max} \int_0^N u(Y_D(l))d\hat{F}^*_{n,K}(l).$$

These estimators are related to $\hat{\theta}_{m,n}^{(I),1}$ and $\hat{\theta}_{m,n}^{(I),2}$ in the following way: $\lim_{m\to\infty}\hat{\theta}_{m,n}^{(I),1} = \hat{\theta}_n^{(I),1}$ and $\lim_{m\to\infty}\hat{\theta}_{m,n}^{(I),2} = \hat{\theta}_n^{(I),2}$. Thus, they are decision rules for which preference uncertainty is completely removed.

First, we analyze the decision rule $\hat{\theta}_{m,n}^{(I),1}$, which is based on the empirical distribution function $\hat{F}_n$ and approximates the policyholder's utility function by $\tilde{u}_m$. In Appendix B, we show that this decision rule is preference-consistent and, since $\lambda > 0$, has a value of preference information of $c(N)$. Looking at Table 3, we can see that the latter value increases as relative risk aversion increases, from \$0 for $\eta = 0.1$ to \$5,639.40 for $\eta = 5$. This is plausible because the deductible choice $N =$ \$40,000 becomes more harmful for more risk-averse individuals.

The expected costs of $\hat{\theta}_{m,n}^{(I),1}$ for $m = 5$ and $m = 20$ are shown in Table 4, along with the expected costs of $\hat{\theta}_n^{(I),1}$. Regarding the impact of preference uncertainty, we observe substantial effects for $\eta = 2$, $\eta = 3$, and $\eta = 5$. For these risk preferences, the expected costs of $\hat{\theta}_{m,n}^{(I),1}$ are

	Costs by degree of risk aversion						Population costs per capita	
	$\eta = 0.1$	$\eta = 0.5$	$\eta = 1$	$\eta = 2$	$\eta = 3$	$\eta = 5$	$\mathbb{P}_{\mathcal{U}}$	$\mathbb{P}_{\mathcal{U}}^{CE}$
Expected costs for $\hat{\theta}_n^{(I),1}$:								
$n = 5$	2.8	45.77	129.61	398.93	940.92	4,500.3	323.61	1,330
$n = 25$	5.15	33.87	91.78	263.45	597.77	2,724.9	200.84	812.05
$n = 100$	6.27	27.25	62.74	143.82	277.79	966.74	79.88	302.24
Expected costs for $\hat{\theta}_{m,n}^{(I),1}$ for $m = 5$:								
$n = 5$	2.8	45.82	129.92	399.93	943.27	4,517.2	324.77	1,334.8
$n = 25$	5.12	34.43	96.12	293.33	697.66	3,154.0	230.75	983.5
$n = 100$	6.2	28.45	71.9	189.63	411.83	1,423.9	112.85	440.77
Expected costs for $\hat{\theta}_{m,n}^{(I),1}$ for $m = 20$:								
$n = 5$	2.8	45.82	129.92	399.93	943.27	4,517.2	324.77	1,334.8
$n = 25$	5.14	34.09	94.67	282.88	617.78	3,187.8	230.97	939.09
$n = 100$	6.27	27.85	67.23	167.71	289.63	1,485.9	113.59	443.94
Expected costs for $\hat{\theta}_n^{(I),2}$:								
$n = 5$	0.72	44.51	0.16	1.34	1.76	5.09	3.47	5.32
$n = 25$	1.28	26.33	0.16	1.29	1.75	4.87	2.85	4.13
$n = 100$	1.72	17.91	0.21	1.18	1.77	4.77	2.71	3.66
Expected costs for $\hat{\theta}_{m,n}^{(I),2}$ for $m = 5$:								
$n = 5$	0.72	45.2	0.16	1.34	3.33	5.1	3.54	5.53
$n = 25$	1.24	30.54	0.15	1.3	2.48	5.22	3.1	4.60
$n = 100$	1.65	24.77	0.53	1.26	2.27	5.17	3.11	4.32
Expected costs for $\hat{\theta}_{m,n}^{(I),2}$ for $m = 20$:								
$n = 5$	0.72	44.99	0.16	1.34	1.77	5.1	3.5	5.36
$n = 25$	1.28	28.73	0.15	1.3	1.83	5.22	3.01	4.41
$n = 100$	1.7	21.72	0.2	1.26	1.89	5.17	2.94	4.06

Note: This table shows the expected costs of choosing an inappropriate deductible (measured in US$) of the two considered preference-dependent decision rules for different $m, n \in \mathbb{N}$ and different degrees of risk aversion. In addition, expected population costs per capita based on the probability measures $\mathbb{P}_{\mathcal{U}}$ and $\mathbb{P}_{\mathcal{U}}^{CE}$ are presented. The results are based on simulations with 1,000 simulation runs.

Table 4: Expected costs for the two preference-dependent decision rules considered

significantly higher than those for $\hat{\theta}_n^{(I),1}$ for both $m = 5$ and $m = 20$. These cost differences thereby increase with the degree of risk aversion. This suggests that our approximate utility functions have problems describing the more pronounced curvature of the utility function at higher η. To better represent this curvature, one must choose a higher m for greater η to reduce the cost deviation from $\hat{\theta}_n^{(I),1}$ to a small amount. Comparing the expected costs for $m = 5$ and $m = 20$ for η greater than 0.5, the expected costs are lower for $m = 20$ except in the case of $\eta = 5$. For $\eta = 1$, $\eta = 2$, and $\eta = 3$, the expected costs for $m = 20$ are close to the expected costs of $\hat{\theta}_n^{(I),1}$. This indicates that for $m = 20$, we have already obtained a good approximation of the corresponding utility functions. However, for $\eta = 5$, the expected costs increase as m increases from 5 to 20 and are still far from those for $\hat{\theta}_n^{(I),1}$. In this case, we need to increase the number of survey questions m in order to further decrease the expected costs.

Next, we analyze the decision rule $\hat{\theta}^{(I),2}_{m,n}$, which is based on the distribution function $\hat{F}^*_{n,K}$ and exploits the approximate utility function $\tilde{u}_m$. Since the distribution $\hat{F}^*_{n,K}$ smooths extreme outliers of the empirical distribution function, we expect the expected costs to be lower for $\hat{\theta}^{(I),2}_{m,n}$ compared to $\hat{\theta}^{(I),1}_{m,n}$. The expected costs shown in Table 4 confirm this hypothesis. Except for individuals with $\eta = 0.5$, expected costs decrease significantly for all $n \in \mathbb{N}$ analyzed. For example, the expected population costs per capita for the probability measure $\mathbb{P}_{\mathcal{U}}$ for $n = 5$ and $m = 20$ decrease from \$113.59 for $\hat{\theta}^{(I),1}_{m,n}$ to \$2.94 for $\hat{\theta}^{(I),2}_{m,n}$. The driving force of these massive cost reductions is the smoothing of the approximation of the loss distribution caused by the use of $\hat{F}^*_{n,K}$. For $\hat{\theta}^{(I),1}_{m,n}$, loss samples with relatively low or no losses lead to the deductible choice $D = \$40{,}000$, which imposes the highest possible cost on individuals with η greater than 0.1. Using the distribution $\hat{F}^*_{n,K}$ smooths the effect of these distributional outliers, such that the policyholder is less likely to choose $D = \$40{,}000$. In particular for small n, for which $\hat{\kappa}^*_{n,K}$ is relatively high, this effect is relatively strong. Therefore, the cost reductions are higher for smaller n.

Similar to $\hat{\theta}^{(I),1}_{m,n}$, the effect of preference uncertainty on the deductible level chosen by $\hat{\theta}^{(I),2}_{m,n}$ is higher, ceteris paribus, when the policyholder is more risk averse. To overcome this uncertainty, one must set m much higher when risk aversion increases. This is especially the case when risk aversion is extremely high (i.e., $\eta = 5$ in our numerical analysis).

Overall, the decision rule $\hat{\theta}^{(I),2}_{m,n}$ seems to be the most appropriate decision rule analyzed in our numerical illustration. First, it leads to by far the lowest expected population cost per capita of all the decision rules analyzed. Second, $\hat{\theta}^{(I),2}_{m,n}$ produces low costs for all risk types and numbers of observations n. This makes it a good candidate for a widely used decision support tool for policyholders. Today, many tools are available to help policyholders to find an optimal insurance contract, mainly by minimizing a person's expected costs and disregarding preference information (Ericson and Sydnor, 2017). As our numerical analysis demonstrates, such tools are significantly inferior to decision rules that consider both distributional information and preference information. Third, $\hat{\theta}^{(I),2}_{m,n}$ is less susceptible to distributional outliers in the loss distribution because it exploits information obtained from available deductible policies. This makes it a better decision rule than $\hat{\theta}^{(I),1}_{m,n}$, which does not adversely affect population groups with certain risk preferences but which is extremely sensitive to distributional outliers.

6.3 Assessment of Decision Rules for Biased Elicitation Processes

In the final part of our numerical analysis, we want to analyze what impact a bias in the elicitation process for the utility function has on the cost figures. For this purpose, we consider

	Costs by degree of risk aversion						Population costs per capita	
	$\eta = 0.1$	$\eta = 0.5$	$\eta = 1$	$\eta = 2$	$\eta = 3$	$\eta = 5$	$\mathbb{P}_{\mathcal{U}}$	$\mathbb{P}_{\mathcal{U}}^{CE}$
Expected costs for $\hat{\theta}_n^{(I),1}$:								
$n = 5$	3.11	44.6	126.25	388.46	916.19	4,387.5	315.75	1,296.6
$n = 25$	8.61	23.83	63.11	187.0	438.14	2,075.3	155.41	617.57
$n = 100$	12.59	6.55	12.52	32.89	68.61	282.63	30.72	91
Expected costs for $\hat{\theta}_{m,n}^{(I),1}$ for $m = 5$:								
$n = 5$	2.8	45.82	129.92	399.93	943.27	4,517.2	324.77	1,334.8
$n = 25$	5.96	30.68	97.74	266.45	634.26	3,221.5	233.88	948.91
$n = 100$	9.46	14.8	57.06	115.41	269.68	1,339.4	104.16	399.48
Expected costs for $\hat{\theta}_{m,n}^{(I),1}$ for $m = 20$:								
$n = 5$	2.95	45.82	129.92	396.94	943.27	4,517.2	324.84	1,334.7
$n = 25$	7.78	28.21	78.87	217.3	573.14	2,861.1	209.02	843.25
$n = 100$	11.62	11.29	28.52	56.28	196.85	967.7	77.68	288.65
Expected costs for $\hat{\theta}_n^{(I),2}$:								
$n = 5$	11.35	1.1	0.41	0.42	0.47	0.57	8.89	5.42
$n = 25$	12.58	2.05	1.0	0.28	0.22	0.35	9.91	5.98
$n = 100$	13.21	2.54	1.32	0.22	0.1	0.42	10.45	6.32
Expected costs for $\hat{\theta}_{m,n}^{(I),2}$ for $m = 5$:								
$n = 5$	10.09	0.37	0.18	0.49	1.77	5.1	8.17	6.13
$n = 25$	10.1	0.59	0.2	0.5	1.81	5.21	8.2	6.18
$n = 100$	10.3	0.7	0.26	0.53	1.81	5.15	8.36	6.27
Expected costs for $\hat{\theta}_{m,n}^{(I),2}$ for $m = 20$:								
$n = 5$	11.34	0.84	0.26	0.49	1.77	5.1	9.17	6.73
$n = 25$	12.35	0.79	0.23	0.32	1.81	5.21	9.94	7.2
$n = 100$	12.74	0.83	0.28	0.35	1.81	5.15	10.25	7.36

Note: This table shows the expected costs of choosing an inappropriate deductible (measured in US$) of the two considered biased preference-dependent decision rules for different $m, n \in \mathbb{N}$ and different degrees of risk aversion. In addition, expected population costs per capita based on the probability measures $\mathbb{P}_{\mathcal{U}}$ and $\mathbb{P}_{\mathcal{U}}^{CE}$ are presented. We assume that the bias is caused by the distortion function s^{-1} of the correction function s presented in (20). For the parameters γ^+, γ^-, and λ_L, we assume that $\gamma^+ = 0.61$, $\gamma^- = 0.69$, and $\lambda_L = 2.25$, which corresponds to the figures estimated by Tversky and Kahneman (1992). The results are based on simulations with 1,000 simulation runs.

Table 5: Expected costs for biased preference-dependent decision rules

the correction function s from (20). For the parameters γ^+, γ^-, and λ_L, we assume that $\gamma^+ = 0.61$, $\gamma^- = 0.69$, and $\lambda_L = 2.25$, which are the parameter values estimated in Tversky and Kahneman (1992). For these parameter choices, the bias in the PE elicitation process leads to more risk-averse estimates for the utility function, which is consistent with what the literature documents about various elicitation methods (cf., e.g., Hershey et al., 1982; Hershey and Schoemaker, 1985; Bleichrodt et al., 2001). The effects of this bias on expected costs are documented in Table 5.

Compared to the unbiased case, the expected costs are lower for all individuals with a coefficient of relative risk aversion not equal to $\eta = 0.1$. Only for nearly risk-neutral individuals with

$\eta = 0.1$ does the bias lead to higher expected costs. This observation can be explained by the fact that the introduced bias in the elicitation method leads to an overestimation of risk aversion. Higher risk aversion, ceteris paribus, reduces the probability of not purchasing insurance (i.e., choosing D = \$40,000). For $\hat{\theta}^{(I),1}_{m,n}$, this effect decreases the average costs of individuals for whom D = \$40,000 is not the optimal deductible choice. Since it is most expensive for them not to purchase insurance, and since the bias reduces the relative frequency of this choice, their costs are substantially reduced. For $\eta = 0.1$, on the other hand, not purchasing insurance is the optimal choice, so expected costs increase due to the bias in the elicitation process. Overall, these two effects lead to a reduction in the expected population costs per capita. For example, the expected population costs per capita for $\theta^{(I),1}_{m,n}$ for $m = 20$ and $n = 100$ decrease from \$113.59 to \$77.68 under the probability measure $\mathbb{P}_{\mathcal{U}}$.

For $\hat{\theta}^{(I),2}_{m,n}$, the results for the different risk preferences go in a similar direction as in the case of $\hat{\theta}^{(I),1}_{m,n}$. However, the magnitude of the changes in expected costs is much lower. For η greater than 0.5, the changes are negligible. Considering $\eta = 0.5$, the bias leads to significantly lower expected costs compared to the unbiased case. This is because the relative risk aversion $\eta = 0.5$ is close to risk neutrality, which leads to a higher probability of not purchasing insurance for certain realizations of the loss distribution approximation $\hat{F}^*_{n,K}$. Due to the bias in the elicitation process, individuals with $\eta = 0.5$ appear more risk-averse than they actually are, which reduces their probability of not purchasing insurance. Since the decision not to purchase insurance is most costly for them, this effect significantly lowers the expected costs.

For policyholders who prefer not to purchase insurance, the bias leads to significantly higher costs, as is the case with $\eta = 0.1$. Since individuals with such risk preferences represent a large fraction of the total population, and since the cost reductions for individuals with $\eta = 1$, $\eta = 2$, $\eta = 3$, or $\eta = 5$ are comparatively small, the expected population cost per capita increases. For example, the expected population costs per capita for $\theta^{(I),2}_{m,n}$ for $m = 20$ and $n = 100$ increase from \$2.94 for to \$10.25 under the probability measure $\mathbb{P}_{\mathcal{U}}$.

Our results suggest that $\hat{\theta}^{(I),2}_{m,n}$ is still by far the best decision rule in terms of expected costs even if the elicitation method used is biased. However, the bias we analyzed causes an increase in expected costs of more than 100 percent. Therefore, one should apply an appropriate correction mechanism to the derived utility functions to avoid higher expected costs.

7 Summary and Conclusion

When purchasing insurance, policyholders face limited information and cannot directly apply theoretical models to choose an optimal deductible. So far, the academic literature has focused on the randomness of future losses as source of uncertainty. However, another source of uncertainty arises from the fact that a policyholder usually does not know the exact functional forms of her loss distribution and utility function. In the present paper, we develop a model that incorporates this second layer of uncertainty into an expected utility framework that is widely used in the literature on deductible policies. This model builds a valuable theoretical foundation for designing online decision aids for policyholders searching for their optimal deductible choice. We first address the question of how to find an appropriate approximation of the loss distribution. To this end, we draw on the empirical distribution function and derive an approximation from the quotes of available deductible policies. The empirical distribution function converges to the true cumulative distribution function as the sample size increases. For the insurance policy offering-based approximation, we find that it converges to the true loss distribution as we increase the number of deductible policies to cover the entire space of possible loss realizations. While the empirical distribution function is a well-known mathematical concept, the offering-based approximation is one of our key contributions to the literature. Given these two approximations, we discuss how they can be optimally combined to obtain a reasonable approximation of the actual loss distribution. The results show that for small sets of loss realizations, the policyholder gives comparatively high weight to the deductible-based approximation. Since we expect most policyholders not to pay attention to the distributional information from insurance policy offerings, our results suggest that the deductible choice in the real world can be improved by drawing on this source of information. This is confirmed by an extensive numerical illustration of the U.S. homeowners insurance market.

After deriving an approximation of the loss distribution, we discuss how to deal with preference uncertainty. First, we show a way to derive a piecewise linear approximation of the utility function using the probability elicitation (PE) method. We present a stepwise survey design for assessing utility functions which extends the existing literature in two ways. First, it leverages information on the distribution of risk preferences in the overall population. Second, it covers the curvature of a utility function more accurately. In addition, we outline the conditions under which a bias in the elicitation method – a well-documented problem in the literature – can lead to a deviation in the optimal deductible choice.

Next, we introduce different decision rules under limited information. To this end, we define different desirable properties and classify the decision rules accordingly. The two decision rules presented in this paper use additional information about the policyholder's risk preferences and are therefore an appropriate choice for widespread decision support. In particular, when we draw on the combined approximation of the loss distribution and preference information, the expected costs resulting from choosing an inappropriate deductible decrease significantly for such types of decision rules.

Our theoretical results have profound policy implications. There is ample empirical and experimental evidence that a significant proportion of policyholders have problems processing important information in insurance settings (cf., e.g., Schoemaker and Kunreuther, 1979; Loewenstein et al., 2013; Handel and Kolstad, 2015). In addition, other psychological phenomena, such as inertia, lead to deviations from optimal insurance decisions. In particular, individuals in vulnerable population groups, such as elderly and low-income individuals, have been shown to make non-optimal insurance decisions (Hanoch et al., 2009; Bhargava et al., 2017). This, in turn, leads to a general wealth redistribution from people who make poor choices to people who make optimal choices, as shown by Ericson and Sydnor (2017).

Against this background, it is critical to provide policyholders with suitable decision making aids to enable them to make insurance decisions that meet their needs. Our results provide guidance on what information is important in what contexts. These findings can be used to design decision aids that better meet an insured individual's needs than those that primarily minimize the expected costs an individual must pay and disregard preference information. These decision aids can then be used as "smart defaults" in the sense of Johnson et al. (2013). By using such defaults, one reduces the cognitive burden of comparing multiple insurance policies and helps policyholders overcome problems related to psychological phenomena such as inertia.

References

Arrow, K. J. (1963). Uncertainty and the Welfare Economics of Medical Care. *American Economic Review 53*(5), 941–973.

Bhargava, S., G. Loewenstein, and J. Sydnor (2017). Choose to Lose: Health Plan Choices from a Menu with Dominated Option. *Quarterly Journal of Economics 132*(3), 1319–1372.

Bleichrodt, H., J. L. Pinto, and P. P. Wakker (2001). Making Descriptive Use of Prospect Theory to Improve the Prescriptive Use of Expected Utility. *Management Science 47*(11), 1498–1514.

Cohen, A. and L. Einav (2007). Estimating Risk Preferences from Deductible Choice. *American Economic Review 97*(3), 745–788.

Cohen, A. and P. Siegelman (2010). Testing for Adverse Selection in Insurance Markets. *Journal of Risk and insurance 77*(1), 39–84.

Corner, J. L. and C. W. Kirkwood (1991). Decision Analysis Applications in the Operations Research Literature, 1970–1989. *Operations Research 39*(2), 206–219.

Doherty, N. A. and H. Schlesinger (1983). The Optimal Deductible for an Insurance Policy When Initial Wealth is Random. *Journal of Business 56*(4), 555–565.

Ericson, K. M. and A. Starc (2012). Heuristics and Heterogeneity in Health Insurance Exchanges: Evidence from the Massachusetts Connector. *American Economic Review 102*(3), 493–497.

Ericson, K. M. and J. Sydnor (2017). The Questionable Value of Having a Choice of Levels of Health Insurance Coverage. *Journal of Economic Perspectives 31*(4), 51–72.

Feldstein, M. S. (1973). The Welfare Loss of Excess Health Insurance. *Journal of Political Economy 81*(2, Part 1 (Mar. - Apr., 1973)), 251–280.

Gertner, R. (1993). Game Shows and Economic Behavior: Risk-Taking on "Card Sharks". *Quarterly Journal of Economics 108*(2), 507–521.

Gollier, C. (2013). The Economics of Optimal Insurance Design. In *Handbook of Insurance*, pp. 107–122. Springer.

Handel, B. R. and J. T. Kolstad (2015). Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare. *American Economic Review 105*(8), 2449–2500.

Hanoch, Y., T. Rice, J. Cummings, and S. Wood (2009). How Much Choice Is Too Much? The Case of the Medicare Prescription Drug Benefit. *Health Services Research 44*(4), 1157–1168.

Hershey, J. C., H. C. Kunreuther, and P. J. H. Schoemaker (1982). Sources of Bias in Assessment Procedures for Utility Functions. *Management Science 28*(8), 936–954.

Hershey, J. C. and P. J. H. Schoemaker (1985). Probability Versus Certainty Equivalence Methods in Utility Measurement: Are they Equivalent? *Management Science 31*(10), 1213–1231.

Insurance Information Institute (2021). Facts + Statistics: Homeowners and Renters Insurance.

Iyengar, S. S. and E. Kamenica (2010). Choice Proliferation, Simplicity Seeking, and Asset Allocation. *Journal of Public Economics 94*(7-8), 530–539.

Iyengar, S. S. and M. R. Lepper (2000). When Choice is Demotivating: Can One Desire Too Much of a Good Thing? *Journal of Personality and Social Psychology 79*(6), 995–1006.

Johnson, E. J., R. Hassin, T. Baker, A. T. Bajger, and G. Treuer (2013). Can Consumers Make Affordable Care Affordable? The Value of Choice Architecture. *PloS one 8*(12), e81521.

Kahneman, D. and A. Tversky (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica 47*(2), 263–291.

Keeney, R. L. (1982). Decision Analysis: An Overview. *Operations Research 30*(5), 803–838.

Klenke, A. (2013). *Probability Theory: A Comprehensive Course*. Springer Science & Business Media.

LaValle, I. H. (1968). On Cash Equivalents and Information Evaluation in Decisions under Uncertainty. *Journal of the American Statistical Association 63*, 252–276.

Loewenstein, G., J. Y. Friedman, B. McGill, S. Ahmad, S. Linck, S. Sinkula, J. Beshears, J. J. Choi, J. Kolstad, D. Laibson, and Others (2013). Consumers' Misunderstanding of Health Insurance. *Journal of Health Economics 32*(5), 850–862.

Luce, R. D. and H. Raiffa (1989). *Games and Decisions: Introduction and Critical Survey*. Courier Corporation.

Marchau, V. A., W. E. Walker, P. J. Bloemen, and S. W. Popper (2019). *Decision Making under Deep Uncertainty: From Theory to Practice*. Springer Nature.

Merkhofer, M. W. (1977). The Value of Information given Decision Flexibility. *Management Science 23*(7), 716–727.

Metrick, A. (1995). A Natural Experiment in "Jeopardy!". *American Economic Review 85*(1), 240–253.

Moffet, D. (1977). Optimal Deductible and Consumption Theory. *Journal of Risk and Insurance 44*(4), 669–682.

Mossin, J. (1968). Aspects of Rational Insurance Purchasing. *Journal of Political Economy 76*(4, Part 1), 553–568.

Murray, M. L. (1972). Empirical Utility Functions and Insurance Consumption Decisions. *Journal of Risk and Insurance 39*(1), 31–41.

Pratt, J. W. (1964). Risk Aversion in the Large and in the Small. *Econometrica 32*(1-2), 122–136.

Rabin, M. (2000). Risk Aversion and Expected-Utility Theory: A Calibration Theorem. *Econometrica 68*(5), 1281–1292.

Raviv, A. (1979). The Design of an Optimal Insurance Policy. *American Economic Review 69*(1), 84–96.

Schlesinger, H. (1981). The Optimal Level of Deductibility in Insurance Contracts. *Journal of Risk and Insurance 48*(3), 465–481.

Schlesinger, H. (1985). Choosing a Deductible for Insurance Contracts: Best or Worst Insurance Policy? *Journal of Risk and Insurance 52*(3), 522–527.

Schlesinger, H. (2013). The Theory of Insurance Demand. In *Handbook of Insurance*, pp. 167–184. Springer.

Schoemaker, P. J. H. and H. C. Kunreuther (1979). An Experimental Study of Insurance Decisions. *Journal of Risk and Insurance 46*(4), 603–618.

Straub, E. (1988). *Non-Life Insurance Mathematics*. Springer.

Sydnor, J. (2006). Sweating the Small Stuff: The Demand for Low Deductibles in Homeowners Insurance. *Manuscript: University of California, Berkeley*.

Sydnor, J. (2010). (Over)insuring Modest Risks. *American Economic Journal: Applied Economics 2*(4), 177–199.

Tversky, A. and D. Kahneman (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty 5*(4), 297–323.

Van Winssen, K. P. M., R. C. Van Kleef, and W. de Ven (2015). How Profitable Is a Voluntary Deductible in Health Insurance for the Consumer? *Health Policy 119*(5), 688–695.

Van Winssen, K. P. M., R. C. Van Kleef, and W. de Ven (2016). Potential Determinants of Deductible Uptake in Health Insurance: How to Increase Uptake in the Netherlands? *European Journal of Health Economics 17*(9), 1059–1072.

Wakker, P. and D. Deneffe (1996). Eliciting von Neumann-Morgenstern Utilities When Probabilities Are Distorted or Unknown. *Management Science 42*(8), 1131–1150.

Zheng, J. (2020). Optimal Insurance Design under Narrow Framing. *Journal of Economic Behavior & Organization 180*, 596–607.

Appendix A

Proof of Theorem 2.1

Proof. From Schlesinger (2013), we know that, ceteris paribus, an increase in the degree of risk aversion of an individual at all wealth levels leads to a decrease in the optimal deductible level. In what follows, we will use this result to prove (i) and (ii).

(i) Looking at the relative risk aversions η_u and $\eta_{u^{(b)}}$ of u and $u^{(b)}$, we find that

$$\begin{aligned}\eta_u(x) = (-x)\frac{u''(x)}{u'(x)} &= (-x)\frac{\frac{\partial^2 u^{(b)}}{\partial^2 x}(x)}{\frac{\partial u^{(b)}}{\partial x}(x)} + (-x)\frac{\partial u^{(b)}}{\partial x}(x)\frac{s''(u^{(b)}(x)}{s'(u^{(b)}(x)} \\ &\geq (-x)\frac{\frac{\partial^2 u^{(b)}}{\partial^2 x}(x)}{\frac{\partial u^{(b)}}{\partial x}(x)} = \eta_{u^{(b)}}(x)\end{aligned}$$

for all $x \in [W - N, W]$. That is, u has a higher degree of risk aversion than $u^{(b)}$. Using the above result from Schlesinger (2013), it follows that $D_K^* \leq D_K^{(b)}$, as desired.

(ii) Looking at the relative risk aversions η_u and $\eta_{u^{(b)}}$ of u and $u^{(b)}$, we find that

$$\begin{aligned}\eta_u(x) = (-x)\frac{u''(x)}{u'(x)} &= (-x)\frac{\frac{\partial^2 u^{(b)}}{\partial^2 x}(x)}{\frac{\partial u^{(b)}}{\partial x}(x)} + (-x)\frac{\partial u^{(b)}}{\partial x}(x)\frac{s''(u^{(b)}(x)}{s'(u^{(b)}(x)} \\ &\leq (-x)\frac{\frac{\partial^2 u^{(b)}}{\partial^2 x}(x)}{\frac{\partial u^{(b)}}{\partial x}(x)} = \eta_{u^{(b)}}(x)\end{aligned}$$

for all $x \in [W - N, W]$. That is, u has a lower degree of risk aversion than $u^{(b)}$. Using the above result from Schlesinger (2013), it thus follows that $D_K^* \geq D_K^{(b)}$, as desired.

□

Proof of Theorem 3.1

Proof. In the following, we define the left-hand limits of F and $\hat{\hat{F}}_K^{ded}$ by $F(x-) = \lim_{y \uparrow x} F(y)$ and $\hat{\hat{F}}_K^{ded}(x-) = \lim_{y \uparrow x} \hat{\hat{F}}_K^{ded}(y)$. First, we note that

$$\lim_{K \to \infty} \hat{\hat{F}}_K^{ded}(x-) = \lim_{K \to \infty} \hat{\hat{F}}_K^{ded}(x) = 0 = F(x) = F(x-)$$

for $x < 0$ and

$$\lim_{K \to \infty} \hat{\hat{F}}_K^{ded}(x-) = \lim_{K \to \infty} \hat{\hat{F}}_K^{ded}(x) = 1 = F(x) = F(x-)$$

for $x \geq N$ by definition of $\hat{\hat{F}}_K^{ded}$ and F. Since f is continuous on the interval $[0, N]$, it is also Riemann integrable on $[0, N]$, so we obtain

$$\lim_{K \to \infty} \int_0^x \hat{\hat{f}}_K^{ded}(z)dz = \int_0^x f(z)dz$$

for $x \in [0, N]$. Using this equation it follows that

$$\lim_{K \to \infty} \hat{\hat{F}}_k^{ded}(0) = \lim_{K \to \infty} 1 - \int_0^N \hat{\hat{f}}_K^{ded}(z)dz = 1 - \int_0^N f(z)dz = \mathbb{P}(L = 0)$$

as well as

$$\lim_{K \to \infty} \hat{\hat{F}}_K^{ded}(0-) = 0 = \mathbb{P}(L < 0) = F(0-).$$

Next, we consider an arbitrary $x \in (0, N)$. In this case we follow

$$\begin{aligned}
\lim_{K \to \infty} \hat{\hat{F}}_K^{ded}(x-) &= \lim_{K \to \infty} \hat{\hat{F}}_K^{ded}(x) \\
&= \lim_{K \to \infty} \hat{\hat{F}}_K^{ded}(0) + \int_0^x \hat{\hat{f}}_K^{ded}(z)dz \\
&= \mathbb{P}(L = 0) + \int_0^x f(z)dz \\
&= F(x) \\
&= F(x-).
\end{aligned}$$

Now we choose an arbitrary $M \in \mathbb{N}$ and define[20]

$$x_j = \inf \left\{ x \in \mathbb{R} \mid F(x) \geq \frac{j}{M} \right\}$$

for $j \in \{0, ..., M\}$ and

$$R_K = \max_{j \in \{1,...,M-1\}} \left(\left| \hat{\hat{F}}_K^{ded}(x_j) - F(x_j) \right| + \left| \hat{\hat{F}}_K^{ded}(x_j-) - F(x_j-) \right| \right).$$

Then we obtain $\lim_{K \uparrow \infty} R_K = 0$. From the definition of x_j it follows that

$$\hat{\hat{F}}_K^{ded}(x) \leq \hat{\hat{F}}_K^{ded}(x_j-) \leq F(x_j-) + R_K \leq F(x) + R_K + \frac{1}{M}$$

and

$$\hat{\hat{F}}_K^{ded}(x) \geq \hat{\hat{F}}_K^{ded}(x_{j-1}) \geq F(x_{j-1}) - R_K \geq F(x) - R_K - \frac{1}{M}$$

[20]The following steps are the same as in the proof of the Glivenko-Cantelli Theorem (Theorem 5.23) in Klenke (2013).

for $x \in (x_{j-1}, x_j)$. Thus, we obtain

$$\limsup_{K\to\infty} \sup_{x\in\mathbb{R}} \left| \hat{\hat{F}}_K^{ded}(x) - F(x) \right| \leq \frac{1}{M} + \limsup_{K\to\infty} R_K = \frac{1}{M}.$$

When M goes to ∞, equation (5) holds, as desired. $\square$

Proof of Lemma 3.2

Proof. (i) This result follows immediately from the fact that $\lim_{n\to\infty} \bar{L} = \mu$ and $\lim_{n\to\infty} \sigma_n^2 = \sigma^2$ $\mathbb{P}$-almost surely by the law of large numbers.

(ii) This relation must hold because $\lim_{K\to\infty} \mu_K^{ded} = \mu$ and $\lim_{K\to\infty} \sigma_{ded,K}^2 = \sigma^2$ by the construction of the cumulative distribution function $\hat{F}_K^{ded}$ and Theorem 3.1.

(iii) First, we note that $\bar{L} \sim N(\mu, \sigma^2/n)$ for large n according to the central limit theorem. We will use this result to derive both G_1 and G_2.

We now assume that $\mu_K^{ded} < \mu$. Then, for $x < 1$ and n large, we obtain the following:

$$\begin{aligned}
\mathbb{P}\left(\tilde{\kappa}_{n,K} \leq x\right) &= \mathbb{P}\left(\frac{\mu - \mu_K^{ded}}{\mu_k^{ded} - \bar{L}} \leq x - 1\right) \\
&= \mathbb{P}\left(\frac{\mu - \mu_K^{ded}}{x-1} \leq \mu_K^{ded} - \bar{L} \leq 0\right) \\
&= \mathbb{P}\left(\mu_K^{ded} - \bar{L} \leq 0\right) - \mathbb{P}\left(\mu_K^{ded} - \bar{L} \leq \frac{\mu - \mu_K^{ded}}{x-1}\right) \\
&= \left(1 - \Phi\left(\frac{\sqrt{n}\left(\mu_K^{ded} - \mu\right)}{\sigma}\right)\right) - \left(1 - \Phi\left(\frac{x}{x-1}\frac{\sqrt{n}\left(\mu_K^{ded} - \mu\right)}{\sigma}\right)\right) \\
&= \Phi\left(\frac{x}{x-1}\frac{\sqrt{n}\left(\mu_K^{ded} - \mu\right)}{\sigma}\right) - \Phi\left(\frac{\sqrt{n}\left(\mu_K^{ded} - \mu\right)}{\sigma}\right).
\end{aligned}$$

It follows immediately from this result that the cumulative distribution function of $\hat{\kappa}_{n,K}^*$ can be approximated by the function G_1 in (8).

Next, we assume that $\mu_K^{ded} > \mu$. Then, we obtain the following for $x < 1$ and n large:

$$\begin{aligned}
\mathbb{P}\left(\tilde{\kappa}_{n,K} \leq x\right) &= \mathbb{P}\left(\frac{\mu - \mu_K^{ded}}{\mu_k^{ded} - \bar{L}} \leq x - 1\right) \\
&= \mathbb{P}\left(\frac{\mu - \mu_K^{ded}}{x-1} \geq \mu_K^{ded} - \bar{L} \geq 0\right) \\
&= \mathbb{P}\left(\mu_K^{ded} - \bar{L} \leq 0\right) - \mathbb{P}\left(\mu_K^{ded} - \bar{L} \leq \frac{\mu - \mu_K^{ded}}{x-1}\right) \\
&= \left(1 - \Phi\left(\frac{x}{x-1}\frac{\sqrt{n}\left(\mu_K^{ded} - \mu\right)}{\sigma}\right)\right) - \left(1 - \Phi\left(\frac{\sqrt{n}\left(\mu_K^{ded} - \mu\right)}{\sigma}\right)\right)
\end{aligned}$$

$$= \Phi\left(\frac{\sqrt{n}\left(\mu_K^{ded}-\mu\right)}{\sigma}\right) - \Phi\left(\frac{x}{x-1}\frac{\sqrt{n}\left(\mu_K^{ded}-\mu\right)}{\sigma}\right).$$

It follows immediately from this result that the cumulative distribution function of $\hat{\kappa}^*_{n,K}$ can be approximated by the function G_2 in (9).
If $\mu_K^{ded} = \mu$, we obtain

$$\tilde{\kappa}_{n,K} = 1 + \frac{\mu - \mu_K^{ded}}{\mu_K^{ded} - \bar{L}} = 1.$$

So, $\hat{\kappa}^*_{n,K} = 1$, as desired.

□

Proof of Theorem 3.3

Proof. First, we note that $U(D)$ and $\hat{U}_{n,K}(D)$ are concave, which can be proved in the same way as in Schlesinger (2013). For brevity, we omit the proof.
Next, we compute the probabilities $\mathbb{P}(\hat{D}^*_{n,K} = D)$ for $D \in \{D_1, \ldots, D_K\}$. For $\hat{D}^*_{n,K} = D_1$, we get

$$\begin{aligned}\mathbb{P}\left(\hat{D}^*_{n,K} = D_1\right) &= \mathbb{P}\left(\hat{U}_{n,K}(D_1) \geq \hat{U}_{n,K}(D) \text{ for } D \in \{D_2, \ldots, D_K\}\right)\\ &= \mathbb{P}\left(\hat{U}_{n,K}(D_1) \geq \hat{U}_{n,K}(D_2)\right)\\ &= \mathbb{P}\left(\hat{U}_{n,K}(D_2) - \hat{U}_{n,K}(D_1) \leq 0\right),\end{aligned}$$

where the second equality follows from the concavity of $\hat{U}_{n,K}$. Using (10), we follow that

$$\hat{U}_{n,K}(D_2) - \hat{U}_{n,K}(D_1) \sim \mathcal{N}\left(U(D_2) - U(D_1), \frac{1}{n}\left(\Sigma_{1,1} + \Sigma_{2,2} - 2\Sigma_{1,2}\right)\right)$$

holds approximately for large n, so we get

$$\mathbb{P}\left(\hat{D}^*_{n,K} = D_1\right) = \Phi\left(\sqrt{n}\frac{U(D_1) - U(D_2)}{\sqrt{\Sigma_{1,1} + \Sigma_{2,2} - 2\Sigma_{1,2}}}\right)$$

for large n, where Φ is the cumulative distribution function of the standard normal distribution.
If $\hat{D}^*_{n,K} = N$,

$$\begin{aligned}\mathbb{P}\left(\hat{D}^*_{n,K} = N\right) &= \mathbb{P}\left(\hat{U}_{n,K}(N) \geq \hat{U}_{n,K}(D) \text{ for } D \in \{D_1, \ldots, D_K\}\right)\\ &= \mathbb{P}\left(\hat{U}_{n,K}(N) \geq \hat{U}_{n,K}(D_K)\right)\\ &= \mathbb{P}\left(\hat{U}_{n,K}(D_K) - \hat{U}_{n,K}(N) \leq 0\right),\end{aligned}$$

where the second equality again follows from the concavity of $\hat{U}_{n,K}$. Using (10), it follows that

$$\hat{U}_{n,K}\left(D_K\right) - \hat{U}_{n,K}\left(N\right) \sim \mathcal{N}\left(U(D_K) - U(N), \frac{1}{n}\left(\Sigma_{K,K} + \Sigma_{K+1,K+1} - 2\Sigma_{K,K+1}\right)\right)$$

holds approximately for large n. Therefore, we obtain the following approximation for large n:

$$\mathbb{P}\left(\hat{D}^*_{n,K} = N\right) = \Phi\left(\sqrt{n}\frac{U(N) - U(D_K)}{\sqrt{\Sigma_{K,K} + \Sigma_{K+1,K+1} - 2\Sigma_{K,K+1}}}\right),$$

where Φ is again the cumulative distribution function of the standard normal distribution. In the last step, we consider the case $\hat{D}^*_{n,K} = D_k$ for $k \in \{2, \ldots, K-1\}$. Here we get

$$\begin{aligned}\mathbb{P}\left(\hat{D}^*_{n,K} = D_k\right) &= \mathbb{P}\left(\hat{U}_{n,K}\left(D_k\right) \geq \hat{U}_{n,K}\left(D\right) \text{ for } D \in \{D_1, \ldots, D_K, N\} \setminus \{D_k\}\right)\\ &= \mathbb{P}\left(\hat{U}_{n,K}\left(D_k\right) \geq \hat{U}_{n,K}\left(D_{k-1}\right), \hat{U}_{n,K}\left(D_k\right) \geq \hat{U}_{n,K}\left(D_{k+1}\right)\right)\\ &= \mathbb{P}\left(\hat{U}_{n,K}\left(D_{k-1}\right) - \hat{U}_{n,K}\left(D_k\right) \leq 0, \hat{U}_{n,K}\left(D_{k+1}\right) - \hat{U}_{n,K}\left(D_k\right) \leq 0\right),\end{aligned}$$

where the second equality again follows from the concavity of $\hat{U}_{n,K}$. Using (10), we follow that

$$\begin{pmatrix}\hat{U}_{n,K}\left(D_{k-1}\right) - \hat{U}_{n,K}\left(D_k\right)\\ \hat{U}_{n,K}\left(D_{k+1}\right) - \hat{U}_{n,K}\left(D_k\right)\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}U\left(D_{k-1}\right) - U\left(D_k\right)\\ U\left(D_{k+1}\right) - U\left(D_k\right)\end{pmatrix}, \frac{1}{n}B^{(k)}\Sigma(B^{(k)})^T\right),$$

where the elements of the $2 \times (K+1)$ matrix $B^{(k)}$ are given by

$$B^{(k)}_{i,j} = \begin{cases}0 & \text{if } j \neq \{k-1, k, k+1\} \text{ or } (i,j) \in \{(1,k+1),(2,k-1)\}\\ 1 & \text{if } (i,j) \in \{(1,k-1),(2,k+1)\}\\ -1 & \text{if } j = k\end{cases},$$

holds approximately for large n. Now we define $\tilde{\Sigma}_k = B^{(k)}\Sigma(B^{(k)})^T$, so that we obtain

$$\mathbb{P}\left(\hat{D}^*_{n,K} = D_k\right) = \Phi_2\left(\sqrt{n}\tilde{\Sigma}_k^{-\frac{1}{2}}\begin{pmatrix}U(D_k) - U(D_{k-1})\\ U(D_k) - U(D_{k+1})\end{pmatrix}\right),$$

where Φ_2 denotes the cumulative distribution function of the two-dimensional standard normal distribution. □

Proof of Theorem 4.1

Proof. First, we note that $U_m(D)$ and $\hat{U}_{m,n,K}(D)$ are concave, which can be proved in the same way as in Schlesinger (2013). For brevity, we omit the proof of this property.

Next, we compute the probabilities $\mathbb{P}(\hat{D}^*_{m,n,K} = D)$ for $D \in \{D_1, \ldots, D_K, N\}$. For $\hat{D}^*_{m,n,K} = D_1$, we get

$$\begin{aligned}
\mathbb{P}\left(\hat{D}^*_{m,n,K} = D_1\right) &= \mathbb{P}\left(\hat{U}_{m,n,K}(D_1) \geq \hat{U}_{m,n,K}(D) \text{ for } D \in \{D_2, \ldots, D_K, N\}\right) \\
&= \mathbb{P}\left(\hat{U}_{m,n,K}(D_1) \geq \hat{U}_{m,n,K}(D_2)\right) \\
&= \mathbb{P}\left(\hat{U}_{m,n,K}(D_2) - \hat{U}_{m,n,K}(D_1) \leq 0\right),
\end{aligned}$$

where the second equality follows from the concavity of $\hat{U}_{m,n,K}$. Using (17), we follow that

$$\hat{U}_{m,n,K}(D_2) - \hat{U}_{m,n,K}(D_1) \sim \mathcal{N}\left(U_m(D_2) - U_m(D_1), \frac{1}{n}\left(\Sigma^{(m)}_{1,1} + \Sigma^{(m)}_{2,2} - 2\Sigma^{(m)}_{1,2}\right)\right)$$

holds approximately for large n, so we get

$$\mathbb{P}\left(\hat{D}^*_{m,n,K} = D_1\right) = \Phi\left(\sqrt{n}\frac{U_m(D_1) - U_m(D_2)}{\sqrt{\Sigma^{(m)}_{1,1} + \Sigma^{(m)}_{2,2} - 2\Sigma^{(m)}_{1,2}}}\right)$$

for large n, where Φ is the cumulative distribution function of the standard normal distribution. If $\hat{D}^*_{m,n,K} = N$,

$$\begin{aligned}
\mathbb{P}\left(\hat{D}^*_{m,n,K} = N\right) &= \mathbb{P}\left(\hat{U}_{m,n,K}(N) \geq \hat{U}_{m,n,K}(D) \text{ for } D \in \{D_1, \ldots, D_K\}\right) \\
&= \mathbb{P}\left(\hat{U}_{m,n,K}(N) \geq \hat{U}_{m,n,K}(D_K)\right) \\
&= \mathbb{P}\left(\hat{U}_{m,n,K}(D_K) - \hat{U}_{m,n,K}(N) \leq 0\right),
\end{aligned}$$

where the second equality again follows from the concavity of $\hat{U}_{m,n,K}$. Using (17), it follows that

$$\hat{U}_{m,n,K}(D_K) - \hat{U}_{m,n,K}(N) \sim \mathcal{N}\left(U_m(D_K) - U_m(N), \sigma^2_K\right)$$

for large n, where $\sigma^2_K = \frac{1}{n}(\Sigma^{(m)}_{K,K} + \Sigma^{(m)}_{K+1,K+1} - 2\Sigma^{(m)}_{K,K+1})$. This gives the following approximation for large n:

$$\mathbb{P}\left(\hat{D}^*_{m,n,K} = N\right) = \Phi\left(\sqrt{n}\frac{U_m(N) - U_m(D_K)}{\sqrt{\Sigma^{(m)}_{K,K} + \Sigma^{(m)}_{K+1,K+1} - 2\Sigma^{(m)}_{K,K+1}}}\right),$$

where Φ is again the cumulative distribution function of the standard normal distribution. In the last step, we consider the case $\hat{D}^*_{m,n,K} = D_k$ for $k \in \{2, \ldots, K\}$. Here we get

$$\begin{aligned}\mathbb{P}\left(\hat{D}^*_{m,n,K} = D_k\right) &= \mathbb{P}\left(\hat{U}_{m,n,K}\left(D_k\right) \geq \hat{U}_{m,n,K}\left(D\right) \text{ for } D \in \{D_1, \ldots, D_K, N\} \setminus \{D_k\}\right) \\ &= \mathbb{P}\left(\hat{U}_{m,n,K}\left(D_k\right) \geq \hat{U}_{m,n,K}\left(D_{k-1}\right), \hat{U}_{m,n,K}\left(D_k\right) \geq \hat{U}_{m,n,K}\left(D_{k+1}\right)\right),\end{aligned}$$

where the second equality again follows from the concavity of $\hat{U}_{m,n,K}$. Using (17), we follow that for large n,

$$\begin{pmatrix}\hat{U}_{m,n,K}\left(D_{k-1}\right) - \hat{U}_{m,n,K}\left(D_k\right) \\ \hat{U}_{m,n,K}\left(D_{k+1}\right) - \hat{U}_{m,n,K}\left(D_k\right)\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}U_m\left(D_{k-1}\right) - U_m\left(D_k\right) \\ U_m\left(D_{k+1}\right) - U_m\left(D_k\right)\end{pmatrix}, \frac{1}{n}B^{(k)}\Sigma^{(m)}(B^{(k)})^T\right),$$

where the elements of the $2 \times (K+1)$ matrix $B^{(k)}$ are given by

$$B^{(k)}_{i,j} = \begin{cases} 0 & \text{if } j \neq \{k-1, k, k+1\} \text{ or } (i,j) \in \{(1,k+1),(2,k-1)\} \\ 1 & \text{if } (i,j) \in \{(1,k-1),(2,k+1)\} \\ -1 & \text{if } j = k \end{cases}.$$

Now we define $\tilde{\Sigma}^{(m)}_k = B^{(k)}\Sigma^{(m)}(B^{(k)})^T$ such that we obtain

$$\mathbb{P}\left(\hat{D}^*_{m,n,K} = D_k\right) = \Phi_2\left(\sqrt{n}(\tilde{\Sigma}^{(m)}_k)^{-\frac{1}{2}}\begin{pmatrix}U_m(D_k) - U_m(D_{k-1}) \\ U_m(D_k) - U_m(D_{k+1})\end{pmatrix}\right),$$

where Φ_2 denotes the cumulative distribution function of the two-dimensional standard normal distribution. □

Appendix B

Preference-dependent Decision Rules

Properties of the Laplace Decision Rule

From Section 4, we know that $\tilde{u}_m$ converges to u for $m \to \infty$. Thus, using the law of large numbers, we obtain

$$\begin{aligned}
\lim_{m,n\to\infty} \hat{\theta}^{(I),1}_{m,n} &= \operatorname*{arg\,max}_{D\in\{D_1,\ldots,D_K,N\}} \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \lim_{m\to\infty} \tilde{u}_m(Y_D(L_i)) \\
&= \operatorname*{arg\,max}_{D\in\{D_1,\ldots,D_K,N\}} \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} u(Y_D(L_i)) \\
&= \operatorname*{arg\,max}_{D\in\{D_1,\ldots,D_K,N\}} \mathbb{E}[u(Y_D(L))] \\
&= D^*_K.
\end{aligned}$$

So, $\hat{\theta}^{(I),1}_{m,n}$ is consistent. Moreover,

$$\lim_{m\to\infty} \mathbb{E}[\hat{\theta}^{(I),1}_{m,n}] = \mathbb{E}\left[\operatorname*{arg\,max}_{D\in\{D_1,\ldots,D_K,N\}} \frac{1}{n}\sum_{i=1}^{n} u(Y_D(L))\right].$$

In general, we cannot interchange the expectation operator and the arg max operator without changing the result. For this reason, $\lim_{m\to\infty} \mathbb{E}[\hat{\theta}^{(I),1}_{m,n}]$ and D^*_K do not coincide in general. That is, $\hat{\theta}^{(I),1}_{m,n}$ is not unbiased in general. Finally, we check whether $\hat{\theta}^{(I),1}_{m,n}$ is preference-consistent. Since $\hat{\theta}^{(I),1}_{m,n}$ is consistent, $\lim_{m,n\to\infty} \mathbb{E}[c(\hat{\theta}^{(I),1}_{m,n}] = 0$. Before we determine the limit $\lim_{n\to\infty} \mathbb{E}[c(\hat{\theta}^{(I),1}_{0,n})]$, we consider the limit

$$\lim_{n\to\infty} \operatorname*{arg\,max}_{D\in\{D_1,\ldots,D_K,N\}} \frac{1}{n}\sum_{i=1}^{n} \tilde{u}_0(Y_D(L_i)).$$

For this limit, using the law of large numbers for $\lambda > 0$, we obtain

$$\begin{aligned}
\lim_{n\to\infty} \underset{D\in\{D_1,\ldots,D_K,N\}}{\arg\max} \frac{1}{n}\sum_{i=1}^{n} \tilde{u}_0(Y_D(L_i)) &= \underset{D\in\{D_1,\ldots,D_K,N\}}{\arg\max} \mathbb{E}[\tilde{u}_0(Y_D(L))] \\
&= \underset{D\in\{D_1,\ldots,D_K,N\}}{\arg\max} \mathbb{E}\left[\frac{1}{N}\left(Y_D(L) - (W-N)\right)\right] \\
&= \underset{D\in\{D_1,\ldots,D_K,N\}}{\arg\max} \mathbb{E}\left[\frac{1}{N}\left(N - \min(L,D) - R(D)\right)\right] \\
&= \underset{D\in\{D_1,\ldots,D_K,N\}}{\arg\min} \mathbb{E}[\min(L,D)] + R(D) \\
&= N
\end{aligned}$$

$\mathbb{P}$-almost surely. So, if $u'' < 0$ and $D_K^* \neq N$, we get

$$\lim_{n\to\infty} \mathbb{E}[c(\hat{\theta}_{0,n}^{(I),1})] = c(N) > 0.$$

That is, $\hat{\theta}_{m,n}^{(I),1}$ is preference-consistent and the corresponding value of preference information is given by $c(N)$. The value of the preference information varies with the degree of risk aversion of the policyholder. In Section 6, we elaborate on this issue by analyzing the value of preference information of various utility functions with constant relative risk aversion (CRRA) for a simulation motivated by the U.S. homeowners insurance market.

Properties of the Optimal Deductible Decision Rule

From Section 4 we know that $\tilde{u}_m$ converges to u for $m \to \infty$. Using the law of large numbers and Lemma 3.2 (i), we obtain

$$\begin{aligned}
\lim_{m,n\to\infty} \hat{\theta}_{m,n}^{(I),2} &= \underset{D\in\{D_1,\ldots,D_K,N\}}{\arg\max} \mathbb{E}[u(Y_D(L))] \\
&= D_K^*.
\end{aligned}$$

So, $\hat{\theta}_{m,n}^{(I),2}$ is consistent. Also,

$$\lim_{m\to\infty} \mathbb{E}[\hat{\theta}_{m,n}^{(I),2}] = \mathbb{E}\left[\underset{D\in\{D_1,\ldots,D_K,N\}}{\arg\max} \int_0^N u(Y_D(l))d\hat{F}_{n,K}^*(l)\right].$$

For large $n \in \mathbb{N}$, the decision rules $\hat{\theta}_{m,n}^{(I),2}$ and $\hat{\theta}_{m,n}^{(I),1}$ approximately coincide, since $\hat{F}_{n,K}^*$ converges more and more to the empirical distribution function $\hat{F}_n$ for large $n \in \mathbb{N}$. Above we already justified why $\hat{\theta}^{(I),1}_{m,n}$ is not unbiased. Using the same reasoning, we follow that $\hat{\theta}^{(I),2}_{m,n}$ is not unbiased. Moreover, using the same arguments as for $\hat{\theta}^{(I),1}_{m,n}$, we follow that

$$\lim_{n\to\infty} \mathbb{E}[c(\hat{\theta}^{(I),2}_{0,n})] - \lim_{m,n\to\infty} \mathbb{E}[c(\hat{\theta}^{(I),2}_{m,n})] = c(N) > 0$$

if $u'' < 0$ and $D^*_K \neq N$. Thus, the value of the preference information is the same as for the decision rule $\hat{\theta}^{(I),1}_{m,n}$.

Appendix C

Parameter Choice for the Correction Function

In our numerical illustration of rule-based decision making in Section 6, we will assume a particular functional form for s introduced by Bleichrodt et al. (2001). To evaluate the hypothetical decision shown in Figure 1, we will assume that the decision maker has preferences that are a combination of prospect theory and expected utility theory. For action a_1, the utility is given by $u(x)$, while the utility of action a_2 is given by

$$u(x) + w^+(p)(u(W) - u(x)) - \lambda_L w^-(1-p)(u(x) - u(W-N)),$$

where w^+ and w^- are the probability weighting functions introduced by Tversky and Kahneman (1992) – that is,

$$w^+ = \frac{\gamma^+}{\left(p^{\gamma^+} + (1-p)^{\gamma^+}\right)^{\frac{1}{\gamma^+}}},$$
$$w^- = \frac{\gamma^-}{\left(p^{\gamma^-} + (1-p)^{\gamma^-}\right)^{\frac{1}{\gamma^-}}},$$

and λ_L is the loss aversion parameter in the sense of Kahneman and Tversky (1979) and Tversky and Kahneman (1992). Under these assumptions, Bleichrodt et al. (2001) derive the correction function

$$s(y) = \frac{w^+(y)}{w^+(y) + \lambda_L w^-(1-y)}. \tag{20}$$

They use this function to correct the utility functions obtained from the PE elicitation procedure. For the parameters estimated in Tversky and Kahneman (1992), that is, $\gamma^+ = 0.61$, $\gamma^- = 0.69$, and $\lambda_L = 2.25$, they document experimental evidence that this correction mitigates the difference with respect to estimates from the CE method. If one additionally uses a specific correction for the CE method, which we do not discuss here, the differences between the two methods even disappear completely.

In general, the correction function presented in (20) neither satisfies $s'' < 0$ nor $s'' > 0$, so one cannot draw general conclusions about the relationship of D_K^* and $D_K^{(b)}$. However, for the parameters assumed in Tversky and Kahneman (1992), $s'' > 0$ holds for the majority of the values $y \in [0,1]$. Thus, we conjecture that $D_K^* \geq D_K^{(b)}$ is likely to hold for this parameter choice. This presumption is confirmed in our numerical illustration in Section 6.