Estimation of (Consumer) Sustainability Benefits from Horizontal Agreements: A Quasi Ex-Post Analysis*

Andy Bartels†, Roman Inderst‡, Christoph Wentzien§

June 2023

Abstract

The European Commission's new horizontal guidelines create scope for sustainability agreements and, to calculate efficiencies, suggest the estimation of consumers' willingness-to-pay (WTP) for so-called non-use value. We contrast estimations obtained from a hypothetical choice experiment conducted for the Netherlands' competition authority (the "Chicken of Tomorrow" case) with estimations based on homescan data documenting consumers' real choices between more or less animal friendly meat, ensuring that the respective alternatives are made highly comparable. Still, estimated WTP for the animal friendly alternative is several times higher in the choice experiment, and we use this finding to inform about several pitfalls both when using experimental data and when using data from true purchases.

Keywords: Sustainability agreements; willingness-to-pay; conjoint analysis

* We thank the Netherlands Authority for Consumers and Markets (ACM) and the AiMark foundation for providing us with the respective data.

† Goethe University Frankfurt, a.bartels@em.uni-frankfurt.de

‡ Goethe University Frankfurt, inderst@finance.uni-frankfurt.de

§ Goethe University Frankfurt, wentzien@econ.uni-frankfurt.de

1 Introduction

The new horizontal guidelines of the European Commission, which govern agreements between competitors, explicitly recognize so-called "sustainability agreements" as a new category, and with it related efficiencies.[1] Such benefits are also increasingly recognized by national authorities, such as in the Netherlands' guidelines,[2] or through legal amendments, such as under the recently changed Austrian competition law. Largely, such efficiencies must still be reflected in higher consumer welfare, e.g., to what extent consumers appreciate the thereby achieved higher ecological sustainability or animal welfare.[3] This puts particular weight on an appropriate measurement of such benefits, and the Commission's guidelines explicitly acknowledge a broadening of approaches.

In this article, we are concerned only with the measurement of so-called non-use value, i.e., appreciated benefits that neither derive from a particular (use) functionality nor from indirect (health) benefits.[4] Our empirical focus will be on animal welfare, as our first dataset comes from a hypothetical choice experiment that the Netherlands' competition authority (Netherlands Authority for Consumers and

---

[1] *See* European Commission, *Draft 'Guidelines on the applicability of Article 101 of the Treaty on the Functioning of the European Union to Horizontal Cooperation Agreements'* (2022), available at: https://ec.europa.eu/competition-policy/public-consultations/2022-hbers_en (last accessed February 17, 2023). For a discussion *see* Roman Inderst & Stefan Thomas, *Integrating benefits from sustainability into the competitive assessment – How can we measure them?*, 12(9) J. EUR. COMP. L.&PRAC. 705 (2021). Together with the draft guidelines, the Commission published its expert report, *see* European Commission, *Incorporating sustainability into an effects-analysis of horizontal agreements' (2022)*, expert report written by Roman Inderst, available at: https://ec.europa.eu/competition-policy/system/files/2022-03/kd0722074enn_HBER_sustainability.pdf (last accessed February 17, 2023).

[2] *See* ACM, *ACM Draft Guidelines, Sustainability agreements – Opportunities within competition law* (2021), available at: https://www.acm.nl/sites/default/files/documents/second-draft-version-guidelines-on-sustainability-agreements-oppurtunities-within-competition-law.pdf (last accessed April 5, 2023) on the Dutch ACM initiative on draft guidelines.

[3] Notably the Commission's guidelines adopt a very broad concept of "sustainability," in line with its general policy commitments. Though a consumer welfare perspective is applied, this has been broadened by the concept of "collective benefits," capturing how an individual consumer may be affected by the choice of other consumers in a relevant market. We note also various calls for broadening the objective of competition policy, such as through the inclusion of sustainability as an independent goal. *See*, e.g., Suzanne Kingston, *Integrating environmental protection and EU competition law: Why competition isn't special*, 16 EUR. L.J. 780 (2010) or Simon Holmes, *Climate change, sustainability, and competition law*, 8 J. ANTITRUST ENFORC. 354 (2020). In what follows, we will not touch further on these issues.

[4] For instance, according to Pearce et al. (2006), non-use value refers to a valuation that is not based on one's actual, planned, or possible use, but potentially considers the use by others, *see* David Pearce, Giles Atkinson & Susana Mourato, *Cost-benefit analysis and the environment: Recent developments*, OECD Publishing (2006). We note that in environmental and resource economics, however, special emphasis is given in particular to health-related consequences. In a technical report commissioned by the Netherlands' and Greece's competition authorities, Inderst et al. (2021) provide an overview of the various methods, *see* Roman Inderst, Eftihios Sartzetakis & Anastasios Xepapadeas, *Technical report on sustainability and competition* (2021), available at: https://www.acm.nl/sites/default/files/documents/technical-report-sustainability-and-competition_0.pdf (last accessed May 4, 2023).

Markets, ACM) conducted in 2014, the famous "Chicken of Tomorrow" case.[5] Data from this served to assess consumers' willingness-to-pay (WTP) for higher animal welfare in the broiler chicken industry,[6] where producers and retailers jointly proposed a higher minimum standard (together with a minimum price). The ACM did not grant permission for this cooperation, also as the estimated WTP for the proposed minor improvement was indeed relatively small and notably below the suggested price increase.

The way the choice experiment was carried out, namely by varying various animal welfare attributes over a considerable scale, allows us to also assess WTP for a larger potential improvement and, in particular, for a more sustainable chicken alternative that was subsequently indeed introduced in the market, without further restrictions to competition: the so-called "beter leven" ("better life," BL in the following) animal welfare label.[7] This is where our second dataset comes in, which consists of GfK's homescan data for chicken filet purchases in the Netherlands from 2018 to 2020. Applying the same (mixed logit) estimation method to observed choices between BL and conventional chicken, we obtain a second estimate of consumers' WTP.

By contrasting these two estimates, our objective is to provide antitrust practitioners with valuable insights into the complexities that emerge when measuring consumer welfare in such contexts. In our case, estimates from the hypothetical choice experiment are not only on average ten times larger, but they are also highly inconsistent with observed behavior. Given actual prices, the distribution of estimated WTP should fully crowd out conventional chicken. But we also argue why WTP estimates based on real purchase behavior risk in turn to considerably underestimate benefits in consumer welfare.

As already noted, the choice experiment decomposed animal welfare in various attributes, such as how chickens are kept, fed, and slaughtered. In principle, this allows to measure incremental WTP for each attribute, and with this for different levels of animal welfare. For instance, WTP depends considerably on whether chickens were fully anesthetized. Spelling out these attributes in sufficient detail should have rendered them salient for decisions in the choice experiment, while we doubt whether real-world buyers of conventional chicken are aware of the fact that chickens are not always sufficiently anesthetized when slaughtered. And even if they are in principle informed about this, such information may not be salient

---

[5] The case is described in Machiel Mulder, Sigourney Zomer, Tim Benning & Jorna Leenheer, *Economische effecten van 'kip van morgen': kosten en baten voor consumenten van een collectieve afspraak in de pluimveehouderij*, Authority for Consumers & Markets (2014). An analysis of this data alone is contained in Machiel Mulder & Sigourney Zomer, *Dutch consumers' willingness to pay for broiler welfare*, 20(2) J. APPL. ANIM. WELF. SC. 137 (2017).

[6] An analysis of WTP in the context of sustainability and (animal) welfare has been carried out in various strands of academic literature, mostly based on surveys. *See*, inter alia, Hans S. Solgaard & Yingkui Yang, *Consumers' perception of farmed fish and willingness to pay for fish welfare*, 113(8) BR. FOOD J. 997 (2011); Caiping Zhang, Junfei Bai & Thomas I. Wahl, *Consumers' willingness to pay for traceable pork, milk, and cooking oil in Nanjing, China*, 27(1) FOOD CONTROL 21 (2012); Anastasios Pagiaslis & Athanasios K. Krontalis, *Green consumption behavior antecedents: Environmental concern, knowledge, and beliefs*, 31(5) PSYCHOL. MARK. 335 (2012).

[7] Chicken is used as a synonym for broiler chicken in the following.

when they fill their shopping cart. Generally speaking, different information and context may be of relevance in the choice experiment compared to the shopping experience in the supermarket. As this can profoundly impact the estimated WTP, this raises the key question of which measurement to rely on when calculating consumer welfare and efficiencies. We will discuss how this issue relates to the notion of "reflective WTP," advanced recently in the context of sustainability agreements.[8]

Our main objective in this paper, however, is to make practitioners aware of the possible range of WTP that different data could generate. We discuss how to narrow this range and suggest how to deal with the remaining potential bandwidth.

As for practical guidelines, we discuss how the presence of a "warm-glow" effect together with the decomposition of animal welfare in various attributes may have contributed to the high WTP for a hypothetical BL chicken in the choice experiment. We also discuss other potential reasons, such as a "hypothetical bias" or the absence of incentives, notably as subjects did not have to pay the price of the chosen chicken alternative. Likewise, we discuss why estimating WTP with real purchase data (such as homescan data) encounters methodological difficulties, leading to an underestimation of mean WTP when the new, more sustainable option is not too widespread and when consumers exhibit a great deal of inertia. We also discuss how the relatively narrow band in which prices and, in particular, price differences for real purchases move, provides a severe handicap for the use of real purchase data.

Methodologically, we connect to a large literature on the elicitation of preferences via conjoint analysis, based on (discrete) choice data. Next to contingent valuation methods, which use data from surveys asking subjects directly about their preferences, conjoint analyses are used widely in marketing and marketing science. For our purpose more relevant is the equally widespread adoption in cost-benefit analysis for public policy assessment and in the field of environmental and resource economics. The literature is surely too wide to review, though we will draw on specific insights below. We, thus, confine ourselves to refer to three consecutive OECD reports written on cost-benefit analysis by experts in the resource economics fields, with an emphasis also on methods to use choice data.[9]

We organize the subsequent discussion as follows. Section 2 introduces the two datasets that we use for our analysis: the choice experiment conducted by the Netherlands' competition authority, ACM, and the homescan data for real purchases. Section 3 introduces the standard econometric framework for discrete choice analyses and with it the estimation of WTP. Results are reported and discussed in Section 4. Based

[8] *See* on this concept Roman Inderst & Stefan Thomas, *Reflective willingness to pay: Preferences for sustainable consumption in a consumer welfare analysis*, 17(4) J. COMP. L.&ECON. 848 (2021).

[9] *See* Pearce et al. (2006), *supra* note 4; Giles Atkinson & Susana Mourato, *Cost-benefit analysis and the environment*, 97 OECD Environment Working Papers (2015); Giles Atkinson, Nils A. Braathen, Ben Broom & Susana Mourato, *Cost-benefit analysis and the environment: Further development and policy use*, OECD Publishing (2018).

on this, we offer a range of practical recommendations in Section 5 and conclude. Section 6 provides a supplementary Appendix.

2 Data

2.1 Data and choice alternatives in the experiment

As described in the introduction, we are able to use data of a choice experiment conducted on behalf of the ACM. This was professionally administered with subjects chosen from the representative CentERdata panel.[10] With a restriction to panelists indicating that they consume chicken meat, 1,603 households took part in the experiment. The experiment was conducted in online format in 2014[11] with the aim of then extracting subjects' WTP for higher animal welfare in raising and slaughtering chicken. Participants faced 15 choice situations, where each time they could choose between two chicken alternatives, representing the purchase of 500 g chicken filet, and the outside option of not purchasing. Choices differed in various attributes as described in Table 1.[12]

Table 1: Attributes in the choice experiment

Attribute	Levels
Lifetime in days	40/60/80
Outdoor access available	No/Yes
Living space (chickens per m^2)	10/15/20
Full anesthesia at slaughter	No/Yes
Number of Dutch consumers purchasing this type of chicken	Small/Large
Observance of higher animal welfare standard guaranteed by collective agreement	No/Yes
Observance of higher animal welfare standard guaranteed by governmental legislation	No/Yes
Price for 500 g of chicken meat in Euro	4/5/6/8/12
Outside option (not purchasing chicken)	No/Yes

The first four attributes in Table 1 are relevant for animal welfare and self-explanatory. We describe below how we make use of these attributes to generate a mapping between a particular choice alternative in the experimental setting and the BL chicken observed in the scanner data.[13] Animal welfare and price are of key relevance for our analysis. The WTP will, however, also depend on the remaining attributes:

[10] For more information regarding the CentERdata panel, *see* https://www.website.centerpanel.nl/nl (last accessed May 01, 2023).

[11] CentERdata, *Willingness to pay voor de kip van morgen*, Tilburg: CentERdata (2014).

[12] From these attributes, 90 choice sets were created in total. Choice sets were divided over six versions of the questionnaires, which resulted in 15 choice sets per questionnaire.

[13] We note that the alternative for which the ACM analyzed the subjects' WTP, referred to as the "Chicken of Tomorrow," differs in various attributes from BL.

the (hypothetical) number of other consumers purchasing the respective alternative and who ascertains the respective claims on animal welfare. We will, thus, also include these attributes for the calculation of WTP of a hypothetical conventional and BL chicken.

2.2 Household scanner data and choice alternatives

Household scanner data covers all fast-moving consumer goods (FMCG) purchases of a representative sample of households. Through cooperation with the AiMark foundation we are able to access the GfK consumer panel, which is a professionally managed sample used widely in industry as well as in research (next to comparable panels administered by Kantar or Nielsen). We cover the period from 2018 to 2020, where the panel contained more than 14,000 households.

We are interested in households purchasing fresh chicken filet, to which we simply refer to as (chicken) "meat." For our WTP analysis we only include conventional chicken filet and filet with a one star BL label. We recall that this label is explained in detail in Appendix 6.1. We note that these conventional and one star BL chickens jointly account for more than 94 % of all chicken filet purchases in the full panel.[14] Furthermore, we only consider purchases in the size category between 400 g and 800 g. Our analysis indicates that larger weights, notably of one kg or more, as well as lower weights represent different markets with only limited sustainability. Also, within the chosen weight range, prices standardized per 500 g do not vary significantly (e.g., in terms of a volume discount). See Appendix 6.2 for details.

We further restrict our consideration to purchases from the five largest retailers in the Netherlands as this will facilitate imputations (see below). In the concentrated Dutch retail landscape, these five retailers account for over 75 % of chicken meat purchases in our panel. Finally, we drop households with less than six months of purchase data and less than three reported shopping trips on average per month.[15] Furthermore, we define price boundaries per year, product (BL and conventional chicken), and retailer to identify purchase incidents that differ stark from the majority of purchase incidents that are associated with our definition of chicken meat.[16] Applying these restrictions, our final dataset consists of

[14] The share of chicken filet is approximately 94.20 % in terms of sales value, while the share in terms of sales volume is approximately 94.91 %.

[15] The latter is a common procedure to exclude households who potentially underreport.

[16] We exclude purchase incidents where the standardized price per 500 g of chicken filet is below 85 % of the standardized yearly retailer-product mean price and is not flagged as a promotion. Moreover, we exclude purchase incidents where the respective price is below 60 % of the standardized yearly retailer-product mean price. In particular, we thereby exclude offers with imminent use-by dates. Lastly, we exclude purchase incidents where the respective price is higher than 115 % of the standardized yearly retailer-product mean price. This ensures that we disregard purchase incidents where the product is (possibly) incorrectly specified.

39,209 purchase incidents from 3,925 different panelists. Amongst these purchase incidents, the fraction BL accounts for is approximately 18.91 % across the three years.[17]

Each purchase incident is described by the identity of the retailer, whether the chicken was conventional or BL, and by the price. We follow much of the literature and assume a market-wide choice set, thus, spanning all five retailers and the respective conventional and BL offers.[18] As we are interested in the incremental WTP for BL vs. conventional chicken, we do not need to take into account an outside option (of not purchasing chicken meat).

As at a given retailer and a given day there may not be purchases of both alternatives, we need to impute the missing price value to construct a consumer's choice environment.[19] The necessary price imputation is heavily facilitated by the fact that retailers practice national pricing in the Netherlands.[20] Our imputation strategy is based on weekly retailer-product-specific prices. Hence, in the first step, we estimate the retailer-specific weekly mean price for BL and conventional meat. In the second step, if necessary, the missing price of a certain week is replaced by the price of the closest week with an available mean price.

2.3 Making the choice environments comparable

We note first that both datasets are derived from professionally managed panels that are representative of the Netherlands' population.[21] The key step is now to map the BL and conventional alternatives in the homescan data to the right sets of attributes in the choice experiment. Based on the definition of the Dutch Society for the Protection of Animals ("Dierenbescherming"), conventional chickens are raised within 38.5 days with 18 chickens per m^2 on average, without outdoor access, and are not fully anesthetized when slaughtered. Instead, one star BL chickens are raised within 56 days with

[17] In 2018, the fraction BL accounts for is 15.47 %, 18.36 % in 2019, and 23.46 % in 2020.

[18] *See*, for instance, applied to an analysis of eggs in the German market Marco J. Kotschedoff & Max J. Pachali, *Higher minimum quality standards and redistributive effects on consumer welfare*, 39(1) MARK. SCI. 253 (2020). We note that one retailer offers only BL chicken in the period under consideration.

[19] We do not have access to retailer scanner data, which would provide information on all sales of an individual outlet.

[20] We refer to Brancatelli et al. (2022) and Brancatelli & Inderst (2022) for a detailed analysis on the basis of the full set of FMCG purchases and for a much longer time period. *See* Calogero Brancatelli, Adrian Fritzsche, Roman Inderst & Thomas Otter, *Measuring income and wealth effects on private-label demand with matched administrative data*, 41(3) MARK. SCI. 637 (2022); Calogero Brancatelli & Roman Inderst, *CPG consumption in times of recession: Novel evidence from matched administrative data* (2022).

[21] A direct comparison of the two samples is made difficult not only by different sociodemographic variables, but notably as, first, the GfK panel has the household and not an individual subject as its key object and as, second, GfK uses bracketing and categorical variables for various attributes.

12 chickens per m², with covered outdoor access, and are fully anesthetized when slaughtered.[22] Thus, the difference between BL and conventional results in a longer lifetime of 17.5 days, six fewer chickens per m², outdoor access, and full anesthesia. We utilize these differences as the foundation for calculating the incremental WTP between the hypothetical BL and the hypothetical conventional chicken.

While we have, thus, pinned down the respective animal welfare attributes for the hypothetical BL and conventional alternatives in the choice experiment, recall that alternatives in the choice experiment are described by additional attributes. Given the share of approximately 19 % of considered purchase incidents that is accounted for by BL, we choose the attribute "small number" to describe BL and "large number" to describe the conventional alternative, both referring to the hypothetical number of other consumers purchasing the respective type of chicken meat. Next, as BL is a collective agreement and as the higher animal welfare standards are neither required nor enshrined in some legislation, for BL we choose "collective agreement" and "no legislation."[23]

3 Econometric framework

We employ for each dataset a mixed logit model to estimate incremental WTP for (hypothetical) BL vs. (hypothetical) conventional chicken. As our procedure is standard, we are rather brief. Still, it seems helpful to formally describe the underlying discrete choice setting. Each individual i is supposed to choose the alternative j that maximizes the utility

$$U_{ij} = V_{ij} + \varepsilon_{ij}, \tag{1}$$

consisting of a systematic component V and a random part ε. The former is specified as

$$V_{ij} = \beta_i' x_{ij}, \tag{2}$$

with x_{ij} denoting the attribute vector of alternative j as seen by individual i and β_i' the vector of the respective coefficients. We follow much of the literature and stipulate that the coefficient for the price, β_{price}, is homogeneous and does not vary due to unobserved heterogeneity.[24] If we now compare two alternatives x_j and $\tilde{x}_j$, dropping the subscript i, the incremental WTP for alternative $\tilde{x}_j$ over alternative x_j is given by

$$Incremental\ WTP_i = \frac{\beta_i'}{-\beta_{price}}(\tilde{x}_j - x_j). \tag{3}$$

[22] *See* Dierenbescherming, *A conscious choice: Introduction of the better life label* (2020), available at: https://beterleven.dierenbescherming.nl/app/uploads/sites/2/2020/04/Interactive-PDF-BLK-introduction-EN-20200604-this-document-1.pdf (last accessed May 19, 2023). For details see Table 6 in the Appendix.

[23] Respectively, for conventional chicken we choose "no collective agreement" and "no legislation."

[24] *See* Kenneth Train, *Discrete choice methods with simulation*, Cambridge University Press (2009).

If we hold the price constant when comparing alternatives, we obtain the incremental WTP for each individual i. For the homescan data, we will also keep retailer identity constant, so that it drops out as well. In what follows, we will always report the incremental WTP for (hypothetical) BL vs. (hypothetical) conventional chicken.

Estimation of the coefficients rests on distributional assumptions. The error terms are supposed to be type I extreme value distributed.[25] Regarding the distribution of coefficients, we follow related work and impose a sign constraint on the coefficients for the (vertical) animal welfare attributes in the choice experiment and the BL indicator in the homescan data. Technically, this is achieved by supposing that the respective coefficient is drawn from a lognormal distribution, while all other coefficients without sign constraints are assumed to be normally distributed.

4 Estimation results and interpretation

In the main text, we report only the distributions of the estimated (incremental) WTP for the (hypothetical) BL alternative compared to conventional chicken. Additional information for the estimation of the different coefficients is contained in Appendix 6.3. Recall that the incremental WTP for BL is obtained by adding up the incremental WTP along the different attributes, as in expression (3).

Table 2 reports results for the choice experiment. According to the estimates, the median incremental WTP for the hypothetical BL alternative is 5.88 Euro. The substantially higher mean of approximately 10.26 Euro results from the fact that a non-negligible fraction of subjects has a seemingly excessive WTP.

Table 2: Quantiles for the WTP distribution (in Euro) with choice experiment data

	Quantiles					Mean	Std.Dev.
	5 %	25 %	50 %	75 %	95 %		
BL vs. Conventional	2.83	4.24	5.88	9.26	23.20	10.26	21.94

Table 3 depicts the respective distribution of WTP as obtained with actual purchase data. The difference is striking. Now the mean WTP is only one tenth. And the overwhelming fraction of households have an estimated WTP of only a few cents. We discuss in what follows, why, however, neither of the two estimates should be entirely trusted, as both suffer from deep-rooted methodological problems.

Table 3: Quantiles for the WTP distribution (in Euro) with the homescan data

	Quantiles					Mean	Std.Dev.
	5 %	25 %	50 %	75 %	95 %		
BL vs. Conventional	0.01	0.02	0.03	0.06	1.92	0.91	5.66

[25] A standard reference is Train, *supra* note 24.

4.1 A first closer look on the estimated WTP

Before interpreting the stark differences, we wish to recapitulate the following. Both estimations were conducted with a substantial amount of data generated either from real purchases recorded in a professional household panel or from a hypothetical choice experiment that was conducted on behalf of the Netherlands' competition authority, with subjects drawn from a renowned and also academically widely used panel. Further, we conducted the estimation by using standard, widespread techniques for discrete choice analysis. We acknowledge that in the choice experiment we constructed the (hypothetical) BL and conventional chicken variants based on the respective attributes.

The high WTP in the choice experiment seems excessive. And it is particularly excessive in light of price differences for real purchase data. Table 4 documents, over the whole sample, the distribution of prices for BL and conventional chicken. We observe that the mean prices differ by only one Euro, and that this price gap applies likewise for the different reported quantiles.

Table 4: Quantiles of the distribution of prices (in Euro) in the homescan data

	Quantiles					Mean	Std.Dev.
	5 %	25 %	50 %	75 %	95 %		
BL	4.38	4.94	4.99	5.36	5.89	5.10	0.43
Conventional	3.65	3.95	3.99	4.18	4.79	4.09	0.31

Table 5 reports price increments for BL vs. conventional chicken. The first line reports, based on individual purchases, the maximum difference between the price of BL and that of a conventional product across full choice sets, noting that we consider five different retailers. For the distribution reported in the second line, we have calculated only differences within a given retailer.

Table 5: Quantiles of the distribution of price differences (in Euro) in the homescan data

	Quantiles					Mean	Std.Dev.
	5 %	25 %	50 %	75 %	95 %		
Maximum diff. across choice set	1.25	1.49	1.88	2.13	2.55	1.86	0.40
Difference inside retailers	-0.16	0.72	0.97	1.18	1.62	0.91	0.52

These prices and price differences based on real purchases can now be confronted with the estimated WTP from the choice experiment. Price differences are lower than even the incremental WTP for the lowest 5 % in the distribution of estimates. In other words, if consumers had in fact the preferences estimated in the choice experiment, virtually all should have purchased the BL alternative. The true market share was, however, under 20 %.

Confronted now with the results from the choice experiment, the distribution of estimated WTP from homescan data seems equally unrealistic. At least three quarters of consumers should have an incremental

WTP of only a few cents, despite the pronounced difference in animal welfare along different dimensions.

In what follows, we offer various rationales for these extreme results. These observations also offer insights in how choice experiments should be designed and why the use of real purchase data may fail. Subsequently, we zoom in on the question of context specificity and why with non-use values the quest for a single, "true" WTP is misguided. Despite this, the estimation of WTP remains an important step, and we provide final guidance on how to make use of it in an assessment.

4.2 Scope for overestimation of WTP in the choice experiment

Recall that non-use value refers to the value that a consumer attributes to the consumption of a good or service that is not related to benefits derived from its use. In particular, the respective value is, thus, not grounded in an immediate physical sensation or in a measurable functionality for a practical purpose. For a consumer it is therefore also difficult to compare different products and, in particular, to attribute a monetary value to the respective difference.

In fact, many subjects in the experiment may have been faced with different attributes contributing to animal welfare for the first time. This may add considerable noise to the elicited valuation, though this argument may not yet support a systematic (upwards) bias. We would also not be convinced that subjects inherently place an excessive value on non-price attributes simply as the choice setting is hypothetical and not incentivized, though we cannot exclude this possibility.[26]

An upwards bias may arise from the moral dimension associated with animal welfare. Subjects may derive benefits from being perceived as contributing towards animals' well-being or from having such a self-perception. Even though the experiment was conducted anonymously online, instead of, e.g., in a face-to-face setting, still subjects may be under the impression that their choices are being watched, in contrast to when they would be filling their shopping cart in a large supermarket.

Subjects may also experience a "warm-glow" effect from acting morally, but we would regard this as not leading by itself to an overestimation, which is so for two reasons. First, such an effect may arise equally when they purchase BL products in the shop, albeit we discuss later how this may be counteracted by limited information and attention as well as inertia. Second, for non-use value it makes, in our view, little sense to focus welfare considerations only on the final "consumption value," thereby ignoring the so-called "decision utility" that consumers may derive from their actual choice. A "warm-glow" effect may, however, bias upwards the estimation of the incremental WTP by the decomposition of animal

[26] More precisely, the respective few attributes have been sufficiently described and are not too abstract in nature. It has been found that this reduces any "hypothetical bias" (*see*, e.g., the discussion in Pearce et al. (2006), *supra* note 4). These factors should also contribute towards increasing "respondent efficiency," *see* Atkinson & Mourato, *supra* note 9.

welfare into several attributes. To recall, the steps in the analysis were: Decomposing animal welfare into different attributes, varying them independently in the choice experiment, and then adding-up estimated incremental WTP for each attribute to obtain incremental WTP for, here, BL vs. conventional chicken. A consumer may now experience a similar "warm-glow" independently of whether only one or all four attributes are varied – or if she ultimately chose BL over conventional chicken.[27] The construction of the two hypothetical choices may then risk double counting a "warm-glow" effect. If this is the case, a simple solution is to let subjects choose directly between the two alternatives. Even when an assessed horizontal agreement proposes different alternatives, these may still be sufficiently small in number so as to allow for direct comparisons.

4.3 Scope for underestimation of WTP with (homescan) purchase data

Though we admittedly do not have evidence for this, we would doubt that consumers generally have knowledge about all the dimensions along which BL differs from conventional chicken or, if such knowledge existed, that this information was salient at the time of purchase. In addition, consumers may exhibit considerable inertia in their everyday shopping. We address potential implications of these observations now in more detail.

For this, we start out on a very practical note. We recall from Table 3 that the estimated incremental WTP for BL vs. conventional chicken equals only a few cents for at least 75 % of consumers. Recall as well that while the share of BL increases, it remains below 20 % throughout the three-year time interval. What is more, for the vast majority of households we observe throughout the same choice of either conventional or BL chicken. Now, if the minimum price difference between BL and conventional chicken in the respective choice sets of a given household was, say, one Euro, strictly speaking we would have learnt only that the incremental WTP for BL was under one Euro, provided that the household continuously chose conventional chicken. The applied standard estimation method, however, pushes the estimated incremental WTP towards the feasible lower extreme, which is zero, as we have applied sign constraints for the animal welfare attributes.[28] Not only does this lead to estimates for these households that are most likely too low, but, given that this applies for the vast majority of households, it also considerably reduces the mean of the estimation. We note that a symmetric problem applies also at the

[27] That a subject's "warm-glow" could be relatively independent of the "distance" between alternatives in terms of animal welfare may be deduced from what has been referred to as the potential failure to pass an "adding-up" or "scope test." A prominent example, there applied to contingent valuation, is Peter A. Diamond & Jerry A. Hausman, *Contingent valuation: Is some number better than no number?*, 8(4) J. ECON. PERSPECT. 45 (1994). For a recent authoritative and less critical view, *see* John C. Whitehead, *Plausible Responsiveness to Scope in Contingent Valuation*, 128 ECOL. ECON. 17 (2016).

[28] Admittedly, this simple illustration abstracts from retailer-specific preferences of a given household. We have also applied a Bayesian estimation approach, where coefficients follow some prior distribution. Results are not much different.

"upper end", i.e., for those consumers who choose BL throughout. But the latter is a far lower fraction of households.

While relative prices, as well as absolute price differences, fluctuate, they remain within narrow bounds (see also Table 5). This is by construction different in the experimental context. Moreover, we conjecture that in the hypothetical choice scenario consumers do not exhibit much of a status-quo bias or inertia, simply as the choice context is hypothetical and new to them. This is quite different in the real shopping context. Again, as the market departs from a situation where initially only conventional chicken were sold, though the share of BL had already climbed to 15 % in 2018, such inertia would exert a downwards bias on the estimate of the WTP for BL chicken.

We already note that in the experimental context, subjects obtained detailed information about the four animal welfare attributes. As reported in the appendix, the estimated mean incremental WTP for full anesthesia is more than three Euros and that for outdoor access exceeds two Euros. Both figures already considerably exceed the price increment for BL chicken as observed in the market. We conjecture, though cannot be sure, that consumers in the market were not fully informed or aware of, for instance, the fact that conventional chicken were not fully anesthetized when being slaughtered. It could well be the case that such information would have convinced far more consumers to purchase BL chicken. On the other hand, it could be argued that in the hypothetical context consumers placed "too much" attention on the respective attributes. In our final remarks on our estimation results, we argue why it would be wrong to try to resolve this difference by seeking a single "true" WTP, but why this does not render the estimation less useful.

4.4 Some final reflections on consumer welfare and WTP estimation

Clearly, using an estimated WTP as an approximation of consumer welfare (or its change) is flawed when the consumer actually cares about some features of a product or service (or its production) but is unaware of it. Next, the estimation could also be flawed when circumstances interfered with the consumer's or subject's choice or response that should not, however, alter consumer welfare. For instance, responses and choices may depend on the order in which different questions or attributes appear or on whether they are highlighted.[29] Taking up again the issue of anesthesia, we would claim that the extent to which subjects or consumers are made aware of this difference between BL and conventional chicken does not fall into such a category. Instead, it is part of the context in which a subject or consumer forms her preferences for the non-use value. Put differently, the respective valuation can only be understood within

[29] Bernheim & Rangel (2008, p.159) have termed these "ancillary conditions," defined as follows: *"a feature of the choice environment that may affect behavior, but that is not taken to be a welfare-relevant characteristic of the chosen object."* (Bert D. Bernheim & Antonio Rangel, *Choice-Theoretic Foundations for Behavioral Welfare Economics* in THE FOUNDATIONS OF POSITIVE AND NORMATIVE ECONOMICS: A HANDBOOK, 155 (Andrew Chaplin & Andrew Schotter (eds), Oxford University Press (2008)).

the respective context.[30] A comparison between WTP estimates from different choice contexts can, thus, not be done reasonably by trying to strip-out the respective context and thereby arriving at some "true" WTP.

In preceding work, we have written on this issue in more detail, notably under the heading of "reflective willingness-to-pay" and with a particular focus on antitrust.[31] This focus has led us to draw some boundaries regarding the admissible context. In particular, we noted that in an hypothetical setting, the context should still be such that it describes a consumer's choice, and not that of a concerned citizen who wishes that she and her fellow citizen cared more about animal welfare or other aspects of sustainability. [32] Otherwise, subjects will not merely frame this as a purchase act, but rather as a "vote" on some societal choice, i.e., regarding what not only they but also other consumers should do. In this case, the estimated WTP would have to be discounted. In our example, however, we believe that the choice context was sufficiently framed as a "narrow" act of own consumption.

Once we have potentially taken into account the discussed problems and potential biases of different estimates, e.g., through respective up- and downwards adjustments, still when different estimates are available, we will be left with a respective range of WTP and, thus, (incremental) consumer welfare. In our previous work we have argued that this may open up additional scope for considering sustainability benefits without infringing on consumer welfare and, thus, without the need to align potentially conflicting objectives. After all, also the (potentially adjusted) boundary of estimated WTP represents an expression of consumer choices.

5 Conclusion

We motivated this article by recent developments in antitrust guidelines and law that aim to include sustainability considerations into an assessment of potentially anticompetitive agreements. In particular, we focused on the measurement and integration of efficiencies based on non-use value that consumers may derive from greater sustainability and that, under a consumer welfare standard, may then have to be traded-off against potential price increases. Efficiencies of this type may also be relevant for the

[30] An alternative view is that preferences are each time newly constructed and, thus, not merely revealed, during the specific decision-making process. The notion of such constructed preferences has been made prominent by Paul Slovic, *The Construction of Preference*, 50(5) AM. PSYCHOL. 364 (1995).

[31] *See* Inderst and Thomas (2021), *supra* note 9, and the discussion also in the expert report for the European Commission's horizontal guidelines, *see* European Commission (2022), *supra* note 1.

[32] Various Nobel laureates have made such distinctions, e.g., between "interests" and "values," *see* Kenneth J. Arrow, *Social Choice and Individual Values*, Yale University Press (1963), between "subjective preferences" and "ethical preferences," *see* John C. Harsanyi, *Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility*, 63(4) J. POLIT. ECON. 309 (1955), or between individual vs. social preferences, where the latter would arise from self-commitment of individuals, *see* Amartya K. Sen, *The Impossibility of a Paretian Liberal*, 78(1) J. POLIT. ECON. 152 (1970).

assessment of mergers, other restrictions to competition, or the potential abuse of dominance. Practitioners, including antitrust authorities, are, thus, well advised to invest into developing the necessary knowledge and skills to apply the respective techniques. But they are equally well advised to learn already now from existing evidence. This allows to put future results into context and to avoid potential pitfalls when setting up and interpreting relevant research.

To those working in this field, the "Chicken of Tomorrow" case is important, as there a competition authority has used methods of discrete choice analysis to estimate the incremental non-use value associated with an agreement, precisely with the provision of chicken meat for which chickens were raised and slaughtered under more animal-friendly conditions. We were able to use this data to calculate also the incremental WTP for an even more animal-friendly alternative than that proposed under the assessed horizontal agreement. This allowed us to then map this experimental data into the choice between conventional chicken meat and meat branded "beter leven" (BL) - a label that was later introduced into the market without infringing on competition.

With our estimations we are able to lay out to practitioners' standard methods of discrete choice analysis, highlighting in particular potential shortcomings. We discussed when the use of real purchase data may significantly underestimate consumers' WTP for a newly introduced product with still relatively low market share, as that of BL chicken. And we discussed reasons for why and when estimates derived from hypothetical choice data may overestimate consumers' valuation. But we also discussed why for the considered non-use values, different contexts in which choices are made may prove significant for the outcome and that this ambiguity cannot be resolved in the sense of unearthing one "true" valuation. This should be kept in mind when antitrust authorities or concerned parties set up and interpret the respective research on which they wish to base their claims or decisions.

6 Appendix

6.1 The BL chicken welfare initiative

The Dutch Society for the Protection of Animals ("Dierenbescherming") launched the BL label in 2007. Today, the label includes criteria for, e.g., laying hens, chicken, pigs, cattle, dairy cows, and turkeys with more than 1,860 farms as well as 430 processing firms and retailers participating.[33]

The BL label comprises three quality levels with the number of stars indicating the level of animal welfare that the respective farm takes into account in production. One star describes chicken production in which the main animal welfare problems have been eliminated. The criteria of the BL label, applied to chicken as considered in this paper, are depicted in Table 6.[34] Additionally, production according to conventional standards is depicted for comparability.

Table 6: BL and conventional criteria for chicken production

Attribute	Conventional Production	One Star BL Production
Indoor floor space	No requirement, 18 chickens per m^2 on average	12 chickens per m^2
Covered outdoor area	No	Yes
Outdoor area	No	No
Slow-growing breed	No	Yes
Enrichment material	No	Yes
Daylight	Not compulsory	Compulsory
Slaughter age	35 - 42 days	56 days
Slaughter method	Electric water bath or gas stunning	Gas stunning recommended
Transport to slaughterhouse	Max. 24 hours	Max. 3 hours

[33] *See* Dierenbescherming, *supra* note 22.

[34] Table 6 is based on an overview of the Dierenbescherming, *supra* note 22. Only chicken filet produced under a conventional or one star BL production standard is considered as these are the types of chicken for which the incremental WTP is estimated. In addition, two and three star BL criteria exist. The objective of the Dierenbescherming is to phase out electric water bath stunning in Dutch slaughterhouses to achieve complete anaesthesia at slaughter.

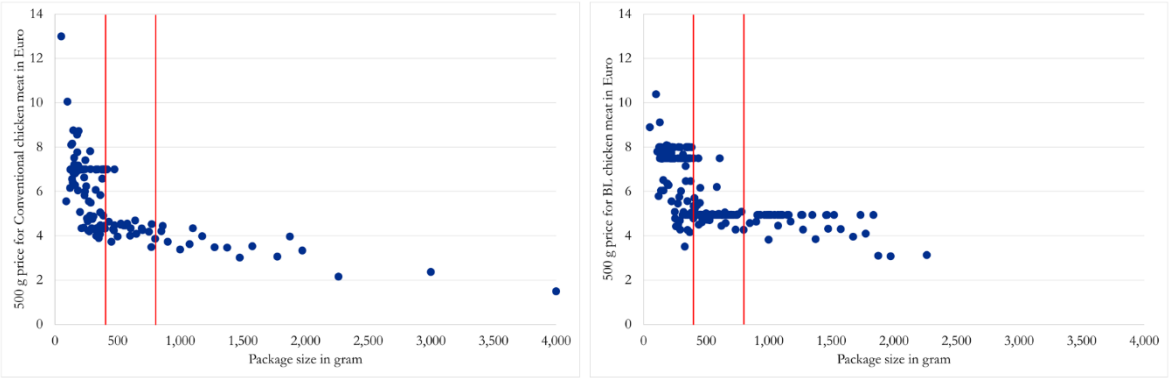
6.2 Choice of purchase incidents

Our object of analysis, the purchase of chicken meat, is not standard for the analysis of homescan data. In fact, research conducted with such data focuses on products in standardized packages, such as coffee, detergents, or canned tuna.

While we make purchase incidents and their prices comparable through normalizing the latter for a 500 g chicken, actual weight varies considerably. In Figure 1, we depict all purchase incidents in terms of weight/volume and normalized prices for BL and conventional chicken in the homescan data. We see that purchases of small volume command a very high price, while there is some discount for particularly large volumes. We stipulate that small, medium, and large volumes are no substitutes, indeed much less so as small and large packages of non- or less perishable products, such as coffee, detergents, but even chocolate. Own experience tells us that it should be a rare scenario that a household purchases one package with a large volume of chicken meat so as to consume from it over several days and meals. Instead, the purchased volume should be determined by the requirements of the respective meal and, thus, notably by the size of the household.

In Figure 1, we have indicated the range of volumes that we consider in our analysis via vertical lines. Over this range, while there is dispersion in prices, we do not observe a significant volume discount.

Figure 1: Package size and price mapping for considered chicken meat



6.3 Estimation details for the choice experiment

Table 7 reports the estimation results for the individual attributes. Recall that the four animal welfare attributes are lognormally distributed, which explains their negative signs.[35] The price coefficient has the expected negative sign and "utility" increases when animal welfare is guaranteed by legislation or by a collective agreement (compared to individual firms' "free market" choice).

Table 7: Mean of attribute coefficients for the choice experiment

Attribute	Coefficient	Std.Err.
Price	-0.406	0.006
Lifetime in days	-5.391	0.220
Outdoor access available	-0.431	0.049
Living space	-4.351	0.170
Anesthesia at slaughter	-0.066	0.037
Number of Dutch consumers	0.274	0.024
Collective agreements	0.273	0.031
Governmental legislation	0.218	0.032
Outside option	-3.994	0.127
Observations	72,135	

Notes: This table depicts the estimation results from the mixed logit model in the choice experiment. Animal welfare attributes are lognormally distributed.

From the underlying distributions for the respective coefficients, Table 8 obtains for each attribute a distribution of the respective incremental WTP in Euro. Note that the values for living space are negative as a higher value of the attribute corresponds to less animal welfare (e.g., more chickens per m^2).

Table 8: WTP distribution (in Euro) for attributes for the choice experiment

	Quantiles					Mean	Std.Dev.
	5 %	25 %	50 %	75 %	95 %		
Lifetime in days	0.00	0.01	0.02	0.03	0.15	0.05	0.25
Outdoor access available	0.76	1.23	1.73	2.50	5.02	2.15	1.63
Living space	-1.80	-0.16	-0.05	-0.03	-0.02	-0.66	3.54
Anesthesia at slaughter	1.05	1.62	2.29	3.86	9.47	3.32	3.09
Number of Dutch consumers	-0.12	0.46	0.72	0.95	1.49	0.70	0.48
Collective agreements	0.65	0.66	0.67	0.68	0.69	0.67	0.01
Governmental legislation	0.48	0.52	0.54	0.56	0.60	0.54	0.04
Outside option	-15.48	-13.83	-11.70	-6.20	-1.42	-9.99	4.82

[35] Consequently, for the subsequent derivation of WTP we need also to transform expression (3).

6.4 Estimation details with homescan data

Table 9 reports the estimation results for the attributes with homescan data. Recall that the BL coefficient is lognormally distributed, which explains its negative sign.[36] As in the choice experiment, the price coefficient has the expected negative sign.

Table 9: Mean of attribute coefficients for homescan data

Attribute	Coefficient	Std.Err.
Price	-2.314	0.020
BL vs. Conventional	-4.607	0.215
Retailer 1	0.745	0.140
Retailer 2	-0.835	0.154
Retailer 4	1.469	0.149
Retailer 5	-2.560	0.194
Observations	352,881	

Notes: This table depicts the estimation results from the mixed logit model in the homescan data. The baseline is defined as retailer 3/conventional.

From the underlying distributions for the respective coefficients, Table 10 obtains for each attribute a distribution of the respective incremental WTP in the homescan data in Euro.

Table 10: WTP distribution (in Euro) for attributes for homescan data

	Quantiles					Mean	Std.Dev.
	5 %	25 %	50 %	75 %	95 %		
BL vs. Conventional	0.01	0.02	0.03	0.06	1.92	0.91	5.66
Retailer 1	-3.10	0.09	0.42	1.46	2.23	0.38	1.53
Retailer 2	-4.05	-0.98	0.07	0.84	2.21	-0.23	1.73
Retailer 4	-3.33	-0.23	1.00	2.35	3.04	0.80	1.87
Retailer 5	-4.77	-1.37	-0.81	-0.45	1.91	-1.01	1.75

[36] Consequently, for the subsequent derivation of WTP we need also to transform expression (3).