Relative versus Absolute Commodity Measurements in Benefit Transfer: Consequences for Validity and Reliability

Robert J. Johnston
George Perkins Marsh Institute and Department of Economics
Clark University
Worcester, MA 01610
rjohnston@clarku.edu

Ewa Zawojska
Faculty of Economic Sciences
University of Warsaw
Warsaw, Poland
ezawojska@wne.uw.edu.pl

Running Head: Benefit Transfer and Commodity Measurement

Corresponding Author: Robert J. Johnston, George Perkins Marsh Institute Clark University, 950 Main St., Worcester, MA 01610. Phone: (508) 751-4619. Email: rjohnston@clarku.edu.

Abstract: Non-market goods can be measured on cardinal or relative scales. Consider a marsh of two hundred acres, of which twenty acres would be affected by a policy. The same affected area can be measured in cardinal terms (twenty acres) or as a relative proportion (ten percent of the marsh). When relative units such as percentages are a scalar transformation of cardinal units, the units of measurement used for modeling are often inconsequential for single-site econometric and welfare analysis. However, this seemingly inconsequential transformation can have significant implications for benefit transfer across sites—a simple observation that remains unacknowledged by the literature. This article provides the first theoretical and empirical evaluation of variable measurement conventions within benefit transfer, deriving conditions under which different types of measurement scales are expected to enhance validity and reliability. Theoretical results are illustrated using an application of discrete choice experiments to coastal flood adaptation in two Connecticut (USA) communities. Empirical findings validate expectations from the theoretical model, suggesting that transfers over goods measured in relative units may substantially outperform transfers over goods measured in cardinal units. These findings imply that the outcome of benefit transfer convergent validity tests may hinge on whether goods are measured in cardinal or relative units.

Funding Acknowledgements: This research is supported in part by the Northeast Sea Grant Consortium, via prime award NA10AOR4170086 to MIT, Sea Grant (Sub-award 5710003190). Opinions do not imply endorsement of the funding agency. The research was undertaken while Ewa Zawojska was a visiting PhD student at Clark University within the Fulbright Foreign Student Program, and was supported by the Foundation for Polish Science (FNP).

JEL Codes: Q51, Q54

Relative versus Absolute Commodity Measurements in Benefit Transfer: Consequences for Validity and Reliability

Abstract

Non-market goods can be measured on cardinal or relative scales. Consider a marsh of two hundred acres, of which twenty acres would be affected by a policy. The same affected area can be measured in cardinal terms (twenty acres) or as a relative proportion (ten percent of the marsh). When relative units such as percentages are a scalar transformation of cardinal units, the units of measurement used for modeling are often inconsequential for single-site econometric and welfare analysis. However, this seemingly inconsequential transformation can have significant implications for benefit transfer across sites—a simple observation that remains unacknowledged by the literature. This article provides the first theoretical and empirical evaluation of variable measurement conventions within benefit transfer, deriving conditions under which different types of measurement scales are expected to enhance validity and reliability. Theoretical results are illustrated using an application of discrete choice experiments to coastal flood adaptation in two Connecticut (USA) communities. Empirical findings validate expectations from the theoretical model, suggesting that transfers over goods measured in relative units may substantially outperform transfers over goods measured in cardinal units. These findings imply that the outcome of benefit transfer convergent validity tests may hinge on whether goods are measured in cardinal or relative units.

Keywords: Benefit Transfer; Measurement Units; Reliability; Stated Preference; Validity; Willingness to Pay

Benefit transfer is defined as the use of pre-existing empirical estimates from primary studies at one or more sites or policy contexts—called study sites—to predict welfare estimates such as willingness to pay (WTP) at other, typically unstudied sites or policy contexts—called policy sites (Johnston and Rosenberger 2010; Johnston et al. 2015). It is widely acknowledged that time, funding and other constraints often lead to a situation in which benefit transfer is the only feasible means to provide economic information required for policy assessment (Griffiths and Wheeler 2005; Iovanna and Griffiths 2006; Johnston, Rolfe and Zawojska 2018). Hence, benefit transfer is a central component of virtually all large-scale benefit-cost analyses in the US, EU and elsewhere (Smith, Van Houtven and Pattanayak 2002; Brouwer and Navrud 2015; Johnston et al. 2015; Loomis 2015; Newbold et al. 2018a).

It is typically assumed that methodological conventions that improve (or detract from) the validity and reliability[1] of original valuation studies will have similar effects on the validity and reliability of benefit transfers based on these original investigations. However, this is not always the case.[2] Consider the measurement scales used to quantify goods or services within an original valuation study. From a purely econometric perspective, a simple linear scaling or normalization of variables within a regression model generally has trivial implications for empirical model results—estimated parameters change by a proportional factor such that the implications of the model are unchanged. For example, travel distance may be measured within a recreation demand model in terms of meters, kilometers or miles, with no substantive effect on model results. Hence, measurement units for modeling are typically determined by researchers as a matter of

[1] For a general discussion of validity and reliability in non-market valuation, see Bishop and Boyle (2019).

[2] For example, Bateman et al. (2011) argue that the addition of variables to improve model performance at an original valuation study site may increase generalization errors when the resulting predictions are used for benefit transfer.

convenience (Boyd et al. 2016). Similarly, most non-market goods can be measured in either cardinal (an absolute quantity) or relative units (Ojea and Loureiro 2011; Johnston et al. 2012; Kling and Phaneuf 2018). Consider a marsh of two hundred acres, of which twenty acres would be affected by a policy. The same affected area may be measured in cardinal terms (twenty acres) or as a relative proportion (ten percent of the marsh). When relative units such as percentages or proportions are a simple scalar transformation of cardinal or absolute units, the units of measurement are generally inconsequential for single-site econometric analysis—at least once the data have been collected.[3] However, as discussed below, seemingly inconsequential transformations of this type can lead to significant consequences for benefit transfer across different sites.

The valuation literature as a whole has given relatively little attention to scaling conventions, such as the use of cardinal versus relative measurement units, reflecting a broader lack of attention to measurement conventions for non-market goods (Johnston et al. 2012; Boyd et al. 2016). A wide array of cardinal and relative measurement conventions appear across the literature (Ojea and Loureiro 2011). For example, a quantity of river habitat accessible to migratory fish may be quantified as either a cardinal number or relative percentage of potential habitat area (Johnston et al. 2012). Similarly, a loss of migratory waterfowl may be quantified in terms of a number of birds or a percentage of the population (Boyle et al. 1994). To the extent

[3] Measurement conventions can sometimes be influential before or during data collection. For example, within stated preference elicitation, survey respondents may react differently to outcomes presented using alternative measurement conventions (Ojea and Loureiro 2011; Johnston et al. 2012). Hence, measurement units for outcomes within stated preference scenarios are often chosen based on survey pretesting. However, once the response data are collected, further scalar variable transformations do not generally affect the results of econometric modeling.

that they appear in the valuation literature, discussions of measurement conventions for non-market goods typically address single-study considerations, often related to scope perceptions within stated preference scenarios (Ojea and Loureiro 2011; see discussion and citations below). Even within this literature, recommendations for cardinal versus relative measurement units are inconsistent. Moreover, once the data are collected, any consistent, ex post linear rescaling of variables (say, between cardinal and relative units) generally has no material impact on site-specific welfare estimates—results for common regression models are scale-invariant in this regard (Gujarati 1988; Babyak 2009).

This lack of attention belies a potentially important impact of measurement conventions on the validity and reliability of benefit transfer that—to the knowledge of the authors—remains unacknowledged. This impact relates to the fact that the econometric invariance of welfare estimates to cardinal versus relative rescaling applies *only when baseline or reference levels (from which relative quantities are calculated) are the same*. If the basis for comparison varies—say across two sites with different baselines for the commodity in question—ex post rescaling between cardinal and relative units is no longer trivial.[4] This challenge is similar to that caused by the comparison of econometric estimates for standardized independent variables (King 1986).

[4] Neoclassical welfare measures for an exogenous commodity change are defined relative to a specified baseline for each individual. Valuation studies often estimate these measures relative to observed or current conditions at a site, which are assumed to be approximately constant for purposes of model estimation. However, baseline conditions can vary, e.g., over time, individuals, or separate areas of a site, and in some cases assumptions of a constant baseline may not be realistic or appropriate. For example, estimation of hedonic property value studies typically relies on observations of different current conditions across homes in different areas. Regardless of these variations, welfare measures are always calculated relative to some specified baseline condition relevant to each individual, and this condition often differs between study and policy sites in a benefit transfer.

Consider an extension of the illustrative example above, wherein twenty acres of marsh would be restored. Assume that coastal Site A has a status quo endowment of two hundred (200) marsh acres, whereas coastal Site B has an endowment of one hundred (100) marsh acres. Hence, an identical twenty-acre restored area represents ten percent of the marsh at Site A, but twenty percent of the marsh at Site B. This also implies that each restored acre represents one-half of a percentage point of the marsh at Site A (0.5% = 1/200), and one percentage point of the marsh at Site B (1% = 1/100). Assume that a primary study at Site A estimates a WTP of $20 per household for the twenty-acre restoration project—equivalent to $1 per restored acre, or $2 per restored percentage point of current marsh. A unit value transfer of this cardinal, *per acre* WTP estimate to Site B would imply an identical $20 value per household for an otherwise identical twenty-acre restoration. In contrast, a transfer of the relative, *per percentage-point* WTP estimate to Site B would imply a per household WTP estimate of $40 for the same restoration. In this simple illustrative example, the transferred value changes by a factor of two, depending on whether one quantifies the commodity using a cardinal or relative scale.

In theory, differences such as these could be at least partially attenuated if one were to implement a transfer using a benefit function that enabled adjustments for variations in status quo baselines across sites. However, the use of benefit functions that enable explicit baseline adjustments remains uncommon in applied benefit transfers. Moreover, illustrative exercises such as that shown above easily confirm that the problem is not (in general) eliminated by the use of typical reduced-form econometric specifications for benefit function transfer which allow WTP to vary as a function of either a commodity baseline or scope of change (e.g., forms that are quadratic in scope to enable diminishing marginal utility).[5] The potential ramifications for validity and reliability are obvious.

In summary, cardinal versus relative scaling decisions that may be inconsequential at a single site can have potentially significant consequences for benefit transfers across different sites. Yet how important are these effects in actual or potential transfer contexts? Moreover, does theory or empirical evidence provide guidance as to whether—and under what conditions—cardinal or relative benefit transfers are likely to be more valid and reliable? (For conciseness, we refer to transfers over commodities measured in cardinal units as "cardinal benefit transfers" and transfers over commodities measured in relative units as "relative benefit transfers.")

Lack of answers to these questions in the literature is not a trivial concern. As implied by the example above, one can easily imagine cases wherein seemingly inconsequential differences in commodity measurement scales could potentially swamp other sources of transfer error. It might also be possible that measurement conventions that are *recommended* for primary studies (not considering benefit transfer possibilities) could lead to systematically *worse* outcomes if those studies are used for benefit transfer. For example, meta-analysis results of Ojea and Loureiro (2011)—suggesting that cardinal measurements are associated with statistically

[5] Many revealed and stated preference valuation studies estimate preference functions that are linear in scope from a single baseline or status quo. The resulting functions are unable to quantify the effect of baseline changes on welfare. In concept, changes in the baseline may be accommodated within meta-analytic benefit functions (e.g., Johnston, Besedin and Holland 2019), but doing so can involve empirical challenges (Newbold et al. 2018b), and baseline changes do not always have statistically significant effects in these models (Johnston, Besedin and Stapler 2017). Estimated benefit functions that are nonlinear in scope (e.g., allowing diminishing marginal utility) can be used to simulate marginal WTP under different baselines. However, reduced-form functions of this type are generally insufficient to guarantee invariance of benefit transfers to the use of cardinal versus relative measurement scales.

significant scope effects—might encourage stated preference survey designers to quantify non-market goods in cardinal units. However, if cardinal measurement conventions were to reduce the accuracy benefit transfers based on those results, it could lead to a tradeoff between primary study measurement error (i.e., a less valid or reliable primary study) and benefit transfer generalization error (i.e., a less valid or reliable transfer).[6]

Despite these concerns, the benefit transfer literature offers no direct insight on this issue. Responding to this lack of guidance, this article provides the first theoretical and empirical evaluation of variable measurement conventions within benefit transfer, focusing on the issue of cardinal versus relative scale. The theoretical model allows for derivation of conditions under which different types of measurement scales are expected to enhance validity and reliability. Illustrative empirical results are provided using an application of discrete choice experiments to coastal flood adaptation in two Connecticut (USA) communities, considering a transfer of values related to coastal asset protection. Empirical results validate expectations from the theoretical model that transfers over goods measured in relative units may often substantially outperform transfers over goods measured in cardinal units. Results from the illustrative case study find that transfer errors are approximately two- to five-times larger, on average, when transfers are conducted over goods measured in cardinal units than in relative units.

Commodity Measurement for Valuation and Benefit Transfer

As noted above, measurement conventions have been given minimal attention within non-market

[6] This tradeoff could be potentially avoided if primary studies were to report sufficient information to enable WTP to be quantified for commodity changes measured in *both* cardinal and relative units. Reporting of this type would enable transfers to be conducted in either type of unit.

valuation (Boyd et al. 2016). One area in which cardinal versus relative measurements have been discussed is the literature on scope effects in stated preference studies (Hanemann 1994; Luisetti, Bateman and Turner 2011; Ojea and Loureiro 2011; Carson 2012; Kling and Phaneuf 2018). The concern in this literature relates to the framing of scope within scenarios and the way in which changes are perceived by respondents. For example, differences that might seem large in cardinal numbers (e.g., 2,000 versus 20,000 birds) might appear trivial when viewed in relative terms (e.g., "much less than 1% of the population" versus "less than 1% of the population"; Desvousges et al. 1993; Hanemann 1994; Carson 2012, p. 34). The recommendations of this literature, however, are not consistent. For example, Luisetti, Bateman and Turner (2011) argue that choice experiment respondents perceive attribute levels in relative rather than absolute or cardinal terms. The meta-analysis of Ojea and Loureiro (2011), in contrast, concludes that a scope test is passed only when goods are "measured in absolute [rather than relative] changes in size" (p. 718). Johnston et al. (2012) take a middle ground, arguing that scenario comprehension is enhanced when changes are presented in both cardinal and relative terms.

Similar lack of consistency is found in the benefit transfer literature. Much of the evidence supporting the validity and reliability of benefit transfer is derived from site-to-site convergent validity tests, in which a benefit transfer estimate is compared to a parallel welfare estimate derived from a primary study at the policy site (Rosenberger and Stanley 2006; Johnston and Rosenberger 2010; Rosenberger 2015).[7] A review of this literature reveals no consistency in cardinal versus relative commodity measurement or scaling conventions.

Conventions vary even for similar commodities and sites. For example, the convergent

[7] For older summaries of these tests, see Bergstrom and De Civita (1999), Brouwer and Spaninks (1999), Rosenberger and Loomis (2003), and Morrison and Bergland (2006), among others.

validity test of Colombo and Hanley (2008) quantifies land use commodities in relative percentages, whereas Johnston and Duke (2008, 2009) describe similar commodities in terms of cardinal land area (acres). Considering water quality, the international benefit transfer evaluation of Czajkowski et al. (2017a) compares WTP for percentage reductions in nutrient loading, and Czajkowski and Ščasný (2010) use a water quality classification relative to average conditions in each country (i.e., both studies employ relative quantities). In contrast, Johnston, Besedin and Stapler (2017), Johnston, Besedin and Holland (2019) and Newbold et al. (2018b) quantify water quality change using a cardinal water quality index calculated from underlying pollutant measures. The benefit transfer test of Morrison et al. (2002) considers river ecology attributes expressed in cardinal units (e.g., square kilometers of wetland), whereas Morrison and Bennett (2004) use similar variables measured in both cardinal and relative units (e.g., percent of river length with healthy vegetation). Similar examples of lack of consistency in measurement conventions can be found when considering other benefit transfers (e.g., for water quality see Hanley et al. 2006; Bateman et al. 2011; Glenk et al. 2015; Brouwer et al. 2016). Few of these evaluations adjust benefit estimates explicitly to accommodate differences in baselines or status quo conditions across sites.

The concern is not solely academic. Benefit transfers conducted to inform governmental benefit-cost analyses or regulatory reviews frequently occur across widely varying baseline conditions, yet do not address the role of commodity measurement conventions. For example, US EPA (2011, Section 8.3) transfers a unit value estimate of WTP per percentage-point increase in fish populations (a relative measure) from a study conducted in a single Rhode Island watershed to policy sites across multiple US states. US EPA and US DOA (2015, Section 9) conduct a unit value transfer of WTP per household per acre for wetland mitigation (a cardinal measure), using the same unit values to approximate benefits across eight different wetland regions in the contiguous US. Both of these studies were conducted for US nationwide regulatory analysis under the Clean Water Act and applied unit value transfers—one using commodities expressed in relative units and one using commodities expressed in cardinal units.

Examples such as these show the diversity of measurement units used for benefit transfer, combined with a tendency to conduct transfers over sites with different baseline conditions. The standard assumption is that measurement conventions were selected by primary study designers based on multiple factors, and reflect appropriate units for each primary study valuation context. Within stated preference studies, for example, measurement conventions are often chosen with input from focus groups and pretests. However, many of the commodities considered in these studies could be potentially rescaled (either ex ante or ex post) between cardinal and relative terms during benefit transfers, and it is not immediately evident which type of measurement would lead to transfers that are more accurate.[8] In particular, it is unclear whether cardinal versus relative commodity measurement conventions chosen based on single-study considerations necessarily lead to studies that are well suited for benefit transfer, or whether the same measurement conventions should be used within benefit transfer itself.

To address these questions, the following section develops a simple theoretical model illustrating the tradeoffs and assumptions that are implied by typical cardinal versus relative transfers. We then provide an empirical illustration of the variations in transfer accuracy that can occur due to rescaling of commodity measurement units alone (between cardinal and relative units), for an otherwise identical benefit function transfer, using choice experiment results.

[8] For example, some stated preference studies communicate non-market goods to respondents using both cardinal and relative units, thereby providing two “native” sets of measurement units that might be used for benefit transfer.

Theoretical Model

To match the empirical application, we orient the theoretical model around the typical random utility framework used as the foundation for most stated preference choice experiments. Assume a simple, illustrative case in which the deterministic (i.e., observable) component of utility for choice alternative p = [1, …, P], with P denoting a number of choice alternatives per valuation question, is represented by a standard linear-in-the-parameters function

(1) $V_{ph} = \gamma_h x_{ph} - \mu_h C_{ph}$,

where subscript h indexes the respondent, x_{ph} is the level of an environmental good measured in cardinal units, C_{ph} is the unavoidable individual cost of the scenario, and γ_h and μ_h are parameters to be estimated. Indexing x_{ph} and C_{ph} by h corresponds to the usual implementation of stated preference surveys, where these characteristics differ as displayed to respondents. Indexing γ_h and μ_h by h accounts for preference heterogeneity across respondents. We make the standard assumptions that $\gamma_h > 0$ and $\mu_h > 0$ (because $\mu_h C_{ph}$ enters utility negatively). The level of the environmental good x_{ph} expresses the amount by which alternative p considers the good to increase or decrease when assessed against a constant baseline or status quo level of the good at the study site, given by some positive level X_0. As common in stated preference surveys, this baseline is assumed not to vary across choice alternatives or respondents at the study site. Although we illustrate the model for a single environmental good, it may easily be extended to the case of multiple goods. Linear functions of this type are ubiquitous in discrete choice experiments, and are generally motivated by the implied (though often unstated) assumption that changes in environmental goods presented in scenarios are sufficiently close to the margin that linear utility functions represent a sufficient first-order approximation to true, nonlinear utility.

Given this specification, the level of good x_{ph} may be rescaled into relative units, such as percentages or proportions, via division by the constant baseline level X_0.[9] We hence define a new (but substantively identical) variable that measures this level in percentage points:

$$(2) \quad \tilde{x}_{ph} = \frac{x_{ph}}{X_0} * 100.$$

For example, x_{ph} might be acres of marsh gained or lost compared to the constant baseline of X_0 acres at the study site, so that $\tilde{x}_{ph}$ is a percentage point gain or loss in the marsh area. For convenience, (2) involves multiplication by the scalar 100 that enables $\tilde{x}_{ph}$ to be interpreted as a percentage on a 100-point scale.[10]

From (2), we obtain $x_{ph} = {}^{X_0\tilde{x}_{ph}}/_{100}$, so that equation (1) may be restated as

$$(3) \quad V_{ij} = \gamma_h \left({}^{X_0\tilde{x}_{ph}}/_{100}\right) - \mu_h C_{ph}.$$

Because ${}^{X_0}/_{100}$ is a constant scalar multiplier, equation (3) may be further restated as

$$(4) \quad V_{ij} = \tilde{\gamma}_h \tilde{x}_{ph} - \mu_h C_{ph},$$

where by definition $\tilde{\gamma}_h = {}^{\gamma_h X_0}/_{100}$.

Following standard estimation approaches for choice experiment data, the parameters in both equations (1) and (4) may be estimated directly using a variety of econometric models for discrete responses, with methods such as mixed logit increasingly prevalent (Train 2009). Equation (1) may be estimated directly if one uses cardinal variable x_{ph} within econometric

---

[9] Rescaling of this type requires a constant baseline across all households, which is a common assumption for stated preference studies.

[10] Alternatively, the level of the good in relative units could also be measured as a proportion (when the multiplication by 100 in (2) is excluded) with no substantive impact on our results.

estimation of the choice model. Equation (4) may be estimated directly if one uses the relative variable $\tilde{x}_{ph}$ within an otherwise identical model. Note that these equations are substantively equivalent, with the resulting coefficient on the good in (4), $\tilde{\gamma}_h = {}^{\gamma_h X_0}/_{100}$, rescaled to account for the scalar transformation from x_{ph} to $\tilde{x}_{ph}$. These are simply two different yet equivalent ways to estimate utility function V_{ph} from the same underlying data.

Marginal WTP, or implicit prices,[11] for the environmental good are likewise consistent at a single site, with WTP per cardinal unit change in the good being

$$(5) \quad WTP_{hx} = \frac{\gamma_h}{\mu_h} > 0,$$

and WTP per relative unit change in the good being

$$(6) \quad \widetilde{WTP}_{h\tilde{x}} = \frac{\tilde{\gamma}_h}{\mu_h} = \frac{\gamma_h}{\mu_h}\frac{X_0}{100} > 0,$$

such that

$$(7) \quad \widetilde{WTP}_{h\tilde{x}} = WTP_{hx}\frac{X_0}{100},$$

and conversely

$$(8) \quad WTP_{hx} = \widetilde{WTP}_{h\tilde{x}}\frac{100}{X_0}.$$

Equations (7) and (8), of course, lead to equivalent WTP estimates for any given change in the environmental good from constant baseline X_0, as the utility function parameters are simply rescaled to account for cardinal versus relative measurement. For example, a one-unit change in

[11] We use the term implicit price as synonymous with marginal WTP following standard conventions in the stated preference literature (Hoyos 2010; Holmes, Adamowicz and Carlsson 2017; Johnston et al. 2017). This term has a slightly different connotation and calculation within the hedonic pricing literature. Nonetheless, under standard assumptions, it may be interpreted similarly as a measure of marginal WTP (Taylor 2017, pp. 265-269).

the cardinally-measured level of the good, that is, $x_{ph} = 1$, valued at WTP_{hx}, is the same as a change in the relatively-measured level of the good equal to $\tilde{x}_{ph} = {}^{1}/_{X_0} * 100$ percentage points valued at $\frac{100}{X_0}\left(\widetilde{WTP}_{h\tilde{x}}\right) = \frac{100}{X_0}\left(WTP_{hx}(\frac{X_0}{100})\right) = WTP_{hx}$, as derived on the basis of (7) and (8). Equations (1) – (8) thus demonstrate that for one site with constant baseline X_0, rescaling between cardinal and relative units is inconsequential for welfare estimation. This, in itself, simply reflects the formal (and well-known) consequence of the invariance of regression results to a scalar variable transformation.

However, now consider a benefit transfer between two *different* sites, A and B, with different baselines for the environmental good, given by X_0^A and X_0^B. As an aside, note that the "close to the margin" assumption that enabled utility to be approximated using a linear function for small changes in x_{ph} may no longer apply when one considers changes from baselines X_0^A and X_0^B, which may differ by non-marginal amounts.[12] In this situation, transfers of WTP for goods measured in cardinal versus relative units imply different assumptions regarding the structure of utility and attendant patterns in WTP. Hence, the transformation is no longer trivial.

First, consider a transfer of per unit WTP for the good measured in *cardinal* units, that is, a transfer of $WTP_{hx} = {}^{\gamma_h}/_{\mu_h}$, from Site A to Site B. Here, one calculates WTP per cardinal unit using estimates from (1) at Site A, and transfers the resulting implicit price to Site B. This cardinal transfer is invariant to the baseline, ${}^{\partial WTP_{hx}}/_{\partial X_0} = 0$, because WTP per cardinal unit is

[12] For example, linear marginal utility might be a reasonable approximation if one considers changes of only a few (e.g., 1, 2, 5) acres from a baseline of 200 acres. However, the marginal utility per acre is likely to change from this level if one starts at a significantly different baseline, such as 20 acres. That is, the marginal utility of "one more" unit will likely differ depending on whether the starting point is 20 or 200 units.

assumed identical across sites. Hence, this transfer does not allow for diminishing marginal utility for changes in *cardinal* units, even when considering baselines that could differ by non-marginal amounts, $X_0^A \neq X_0^B$. A corollary result is that the associated WTP per *relative* unit, $\widetilde{WTP}_{h\tilde{x}} = WTP_{hx}(^{X_0}/_{100})$ as per (7), increases with the baseline (i.e., $^{\partial\, \widetilde{WTP}_{h\tilde{x}}}\big/_{\partial X_0} = {}^{WTP_{hx}}/_{100} > 0$), everything else held constant. From an intuitive perspective, this result implies that individuals are willing to pay more for a given relative change in the level of a good, *ceteris paribus*, when that relative change represents a greater number of cardinal units.

Now consider a transfer of marginal WTP for the good measured in *relative* units, given by $\widetilde{WTP}_{h\tilde{x}} = {}^{\tilde{\gamma}_h}/_{\mu_h}$ as per (6). In this case, the *relative* WTP value is assumed constant across sites. Hence, by equation (8), the marginal utility per *cardinal* unit now decreases with increases in the baseline, $^{\partial WTP_{hx}}/_{\partial X_0} = -100\, \widetilde{WTP}_{h\tilde{x}} X_0^{-2} < 0$, reflecting a specific form of diminishing marginal utility. This transfer of fixed WTP per relative unit implies that WTP per cardinal unit diminishes if one conducts a transfer from Site A to Site B, where $X_0^A < X_0^B$. Intuitively, this is because the value per *relative* unit is assumed constant across sites, and there is a greater number of cardinal units per relative unit when the baseline increases from X_0^A to X_0^B.

These results demonstrate that the use of cardinal versus relative unit value transfers between sites with different baselines implies different assumptions about utility. The degree to which these assumptions hold will influence the validity of the resulting transfers. For unit values, cardinal transfers impose a potentially strong assumption of constant marginal utility per cardinal unit (and, hence, constant WTP_{hx}) across sites, regardless of differences in baseline conditions. This may not be credible if the difference in baselines between sites is large. On the other hand, relative transfers imply a specific mathematical form of diminishing marginal utility per cardinal unit. Which of these assumptions is closer to actual conditions is unknown, but often one might wish to allow for some form of diminishing marginal utility in cardinal units if baselines differ across study and policy sites. For example, the marginal utility per marsh acre might be expected to diminish at sites that already have a large number of marsh acres (e.g., X_0^A = 1000 acres), compared to other sites with few baseline acres (e.g., X_0^B = 10 acres). In such cases, relative unit value transfers should—at least in theory—have greater face validity than cardinal transfers, holding all else constant.[13]

Although straightforward in concept, the issues formalized above have not been acknowledged by benefit transfer literature or accommodated in guidance for transfer methods. There has been no recognition that (a) the variable measurement units alone might have potentially significant effects for the outcomes of benefit transfer validity and reliability tests, and (b) some measurement conventions may be more likely to engender valid and reliable transfers. However, whether and how these theoretical concerns affect transfer validity and reliability in applied settings is an empirical question.

Empirical Application

To illustrate and verify predictions implied by the theoretical model, we use data from stated preference discrete choice experiments (DCEs) implemented in Old Saybrook and Waterford, two coastal communities in Connecticut, USA. The DCEs were developed to elicit residents' preferences for methods of coastal flood adaptation, such as shielding the coastline with hard and

[13] Similar illustrations may be developed for various types of benefit function transfer. Naturally, the structural implications of commodity measurement conventions for these transfers vary across different functional forms assumed for utility.

soft defenses, and outcomes of the adaptation efforts, including protection of homes, coastal marshes and beaches. The two communities are located in the same part of the state but differ in flood risk, which is higher in Old Saybrook. The communities also differ in endowments and vulnerabilities of natural assets considered in the DCE. For example, at the time of data collection in 2014, Old Saybrook had a total of approximately 497 acres of coastal marsh, of which up to 50 acres could be lost within the next decade as predicted by coastal scientists (reflecting potential erosion losses). Waterford, in contrast, had a total of only 77 acres, of which 15 acres could be lost. Differences such as these imply that transformations from cardinal to relative attribute metrics—although inconsequential when analyzing DCE results for each site in isolation—could have non-trivial implications when transferring WTP between the two sites.

Details of the underlying DCEs have been discussed by prior works (Johnston and Abdulrahman 2017; Johnston, Makriyannis and Whelchel 2018; Makriyannis, Johnston and Whelchel 2018); the description here focuses only on elements that are central to the present analysis. The DCEs conducted in Old Saybrook and Waterford are identical beyond the above-noted differences in quantitative attribute levels and baselines, facilitating the proposed validity and reliability tests. The questionnaires were designed and pretested based on an extensive, more than two-year collaboration of economists, environmental scientists, municipal officials and stakeholders, including thirteen focus groups with residents, supplemented with individual cognitive interviews.[14] Verbal protocols (Schkade and Payne 1994) were used within focus groups and interviews to gain insight into respondents' comprehension and decision processes. Development of policy scenarios considered in the DCEs was informed by data from the Center for Climate Systems Research at Columbia University, NASA's Goddard Institute for Space

[14] For a general description of cognitive interview methods, see Kaplowitz, Lupi and Hoehn (2004).

Studies, the Nature Conservancy and the National Oceanic and Atmospheric Administration.

All DCE respondents were presented with a sequence of three choice tasks, each of which provided three options for the flood adaptation policy. In every choice task, respondents were asked to select (vote for) the policy option that best matched their preferences. Two of the three options proposed new flood adaptation strategies, while the third option (henceforth, referred to as status quo) involved no new action. Prior to displaying choice tasks, the questionnaire discussed tradeoffs between approaches to flood protection, presented inundation scenario forecasts for the mid-2020s, and described the future (as of the mid-2020s) status quo assuming no change in adaptation actions. The information was conveyed using text, graphics, maps and photographs, based on input from focus group pretesting and individual cognitive interviews. In line with recommendations for stated preference research (Johnston et al. 2017), the questionnaire also drew respondents' attention to important factors to consider while making choices. For example, the script framed responding in choice tasks as voting decisions, which resembled voting in actual referenda. Second, the information instructed respondents to treat each choice task as an independent choice, and not to compare policy options across different choice tasks. Third, study consequentiality was highlighted via a statement that results would be shared with policy makers. Finally, the questionnaire included a reminder of budget constraints.

Flood adaptation policy options were characterized by five non-monetary attributes and a monetary attribute defined as an unavoidable cost to a household (i.e., an increase in annual taxes and fees). The non-monetary attributes included: (1) the number of homes expected to flood in high intensity storms, (2) the acreage of coastal marshes and (3) of beaches and dunes expected to be lost due to flooding or erosion, (4) the length of hard-armored coastline and (5) the general emphasis of adaptation efforts, specifically whether emphasis would be placed on hard or soft coastal defenses. All outcomes were forecast as of the mid-2020s.

Quantitative attribute levels were presented in both cardinal and relative units. Proposed policy changes in relative units were expressed as percentages with regard to the current level of the respective attribute. Examples include the percentage of homes or marshes in the community that would be flooded or lost. At the same time, scenarios provided information on these attribute levels in cardinal units, such that the attribute levels for homes and marshes specified the number of homes or marsh acres that would be flooded or lost. Table 1 describes the attributes and their levels within the Old Saybrook and Waterford DCEs. Attribute levels corresponded to feasible flood adaptation outcomes for each site, which were identified based on the above-mentioned data sources and consultations with local experts.

Figure 1 shows an example of a choice task. Fractional factorial experimental designs for the DCEs at the two sites were optimized for D-efficiency (Sándor and Wedel 2001, 2002; Ferrini and Scarpa 2007; Scarpa and Rose 2008) for main effects and two-way interactions. This yielded 72 profiles, which were blocked in 24 booklets of three choice tasks each. Blocking minimized correlation between the design variables and assigned blocks.

The survey was implemented from May to June 2014. Questionnaires were distributed to a random sample of Old Saybrook and Waterford households via US mail, with follow-up mailings to increase response rates (Dillman, Smyth and Christian 2009). Of 1,152 mailed questionnaires for the DCE versions analyzed here (576 for each community), 995 were deliverable and 329 were returned, yielding a response rate of 33.1%. These include some booklets returned incomplete, because respondents were instructed to return a blank questionnaire if they did not wish to participate. The resulting net return of complete and usable booklets was 283 (a 28.4% usable response rate), yielding 815 usable choice observations for both DCEs combined (408 for Old Saybrook; 407 for Waterford).

The effect of socio-demographic adjustments on stated preference benefit transfer has been explored previously (e.g., Barton 2002; Johnston and Duke 2010), and is not a primary focus here. Hence, the analysis is conducted using the most common and streamlined format for choice modeling of this type, excluding socio-demographic variables. Nevertheless, because there are some statistically significant differences in socio-demographic characteristics across the samples of Old Saybrook and Waterford respondents,[15] we also estimate a supplementary model including a set of socio-demographic interactions. This model is estimated to evaluate the robustness of our findings, and is presented in Appendix A. As shown in this appendix, adjustments for socio-demographic differences within the model have trivial implications for benefit transfer results and related hypothesis tests, suggesting that model results are robust to the inclusion or exclusion of socio-demographic variables. We therefore present our primary empirical results using the more streamlined choice models that omit these variables.

Econometric Model

Estimation of discrete choice models to analyze preferences is grounded in the standard random utility framework (McFadden 1974; Hanemann 1984). Here, we estimate the model in WTP-space to avoid challenges for welfare assessment associated with the use of preference- (or utility-) space with a random coefficient on policy cost (Train and Weeks 2005; Scarpa, Thiene and Train 2008).

To formalize the model, we assume that utility of household *h* from policy scenario *p*,

[15] These differences are not surprising, because the underlying population characteristics of the two communities are not identical.

$U_{ph}(\cdot)$, is a function of vector $\boldsymbol{X_{ph}}$ of non-monetary characteristics of the policy scenario, as specified by the attribute levels, and the household cost C_{ph} of the policy scenario. Following standard practice, we further assume that utility includes both a systematic component $V_{ph}(\cdot)$, and a random component treated as an unobservable error, $\varepsilon_{ph}(\cdot)$, such that

(9) $U_{ph}(\cdot) = U_{ph}(\boldsymbol{X_{ph}}, C_{ph}) = V_{ph}(\boldsymbol{X_{ph}}, C_{ph}) + \varepsilon_{ph} = \boldsymbol{\gamma_h}'\boldsymbol{X_{ph}} - \mu_h C_{ph} + \varepsilon_{ph}.$

The interpretation of parameters $\boldsymbol{\gamma_h}$ and μ_h is specific to the model type and is detailed below. Multiple types of empirical discrete choice models are consistent with the framework in (9), depending on assumptions regarding the structure of the observed and unobserved utility components and other aspects of choice behavior. Mixed logit modeling is common. This approach enables parameters to vary over households (as indicated by indexing them over h) according to a predefined, multivariate distribution (Train 2009). We follow this approach here.

To further develop the model, we specify $\mu_h = {}^{\alpha_h}/_{\sigma_h}$ as the preference-space coefficient on the policy cost, with α_h denoting the underlying marginal utility of income and σ_h as the logit scale parameter; $\boldsymbol{\gamma_h} = {}^{\boldsymbol{\beta_h}}/_{\sigma_h}$ as a conforming vector of coefficients on non-monetary attributes $\boldsymbol{X_{ph}}$, where $\boldsymbol{\beta_h}$ is a vector of underlying marginal utilities of these attributes; and ε_{ph} as an i.i.d. type-one extreme-value error with constant variance $Var(\varepsilon_{ph}) = {}^{\pi^2}/_{6}$ (Train and Weeks 2005; Scarpa, Thiene and Train 2008). From (9), a vector of parallel WTP coefficients may be calculated as the ratio of coefficients on non-monetary attributes and the coefficient on cost, that is, as $\boldsymbol{\omega_h} = {}^{\boldsymbol{\gamma_h}}/_{\mu_h} = {}^{\boldsymbol{\beta_h}}/_{\alpha_h}$. We can hence rewrite (9) to derive the parallel WTP-space specification of the model (Train and Weeks 2005):

(10) $U_{ph}(\cdot) = \mu_h[(\boldsymbol{\gamma_h}/\mu_h)'\boldsymbol{X_{ph}} - C_{ph}] + \varepsilon_{ph} = \mu_h(\boldsymbol{\omega_h}'\boldsymbol{X_{ph}} - C_{ph}) + \varepsilon_{ph}.$

Equation (10) is behaviorally equivalent to the preference-space specification in (9), but vector $\boldsymbol{\omega_h}$

includes a set of random coefficients representing estimates of WTP for each choice attribute (or implicit prices), assumed to be normally distributed. We further specify $\mu_h = e^{v_h}$, with v_h as the underlying latent normal factor defining the lognormally distributed cost coefficient (Train and Weeks 2005; Scarpa, Thiene and Train 2008). This ensures positive marginal utility of income.[16]

Our emphasis on investigation of benefit transfer validity and reliability implies a set of tests focused on implicit price (marginal WTP) differences between the two sites (assuming that implicit prices from one site are used to approximate parallel implicit prices at the other; Rosenberger and Stanley 2006; Rosenberger 2015). To facilitate these tests, we estimate the model using pooled data from Old Saybrook and Waterford, with (10) extended to allow for systematically varying preferences and scale across the two community samples.[17] This is accomplished following the approach of Czajkowski et al. (2017b), in which we specify the vector of WTP coefficients within the mixed logit model, $\boldsymbol{\omega_h}$, as

[16] Specifications of this type are recommended because cost coefficient distributions that overlap zero lead to undefined WTP moments (Daly, Hess and Train 2012; Johnston et al. 2017). However, to verify whether the assumption of positive marginal utility of income is supported by our data, we estimate a preference-space model with a fixed and unconstrained cost parameter and all non-monetary parameters specified as random and normally distributed. The sign of the cost coefficient in this model confirms a positive marginal utility of income. We further find no statistically significant differences in WTP estimates across this alternative preference-space model with a fixed cost coefficient and the WTP-space model with all random parameters, which is the central model in this paper.

[17] Our primary econometric analysis involves a pooled model for the two communities. The pooled model provides statistical efficiency gains and enables direct tests of WTP differences between communities. However, in most benefit transfer contexts, underlying valuation study data are not accessible and no study data are available for policy sites. Hence, estimation of a pooled model is impossible. Therefore, in Appendix B, we conduct a similar analysis using independent mixed logit models estimated for each community. As shown in the appendix, our results are robust to this alternative estimation approach.

$$(11) \quad \boldsymbol{\omega_h} = \boldsymbol{\omega_h^*} + \boldsymbol{\rho} W_h.$$

Vector $\boldsymbol{\omega_h^*}$ is distributed multivariate normal (correlated random parameters) with means and covariance matrix to be estimated, reflecting the distribution of WTP estimates. The vector $\boldsymbol{\rho}$ is a set of parameters that enables WTP estimates to vary according to binary indicator (dummy) variable W_h that takes a value of one for Waterford observations and zero otherwise. We also redefine

$$(12) \quad \mu_h = e^{(v_h + \tau W_h)},$$

with τ as a coefficient that allows the marginal utility of income and scale to vary systematically between the two split-samples, with v_h as defined above.[18] The resulting main effect estimates for the means of $\boldsymbol{\omega_h^*}$, therefore, represent mean implicit prices for the Old Saybrook sample, whereas estimates for $\boldsymbol{\rho}$ represent a difference in the mean implicit prices between Old Saybrook and Waterford.

The model is estimated via simulated maximum likelihood with 7,000 Sobol draws.[19] Multiple alternative model specifications were evaluated to ensure robustness, and all generated results similar to those presented below, with key findings unchanged. Drawing from mixed logit model results, we evaluate the performance of WTP transfers for cardinally- and relatively-measured outcomes separately, with respect to both validity and reliability. Specifically, the WTP-space model in (10) is estimated using goods measured in relative units, with transformations between relative and cardinal welfare measures then made following equation (8). These cardinal and relative welfare measures are subsequently used for benefit transfer convergent validity tests between Old Saybrook and Waterford.

[18] The parameter μ_h captures variation in both scale and the marginal utility of income (Thiene and Scarpa 2009).

[19] We use a custom code developed in Matlab available at https://github.com/czaj/DCE under CC BY 4.0 license.

Validity testing verifies whether WTP values are statistically identical across the two sites. Thus, we calculate WTP values for attributes measured in both cardinal and relative units for Old Saybrook and for Waterford, and compare which type of the benefit transfer (that is, in cardinal or relative units) more often generates statistically valid WTP transfers. A parallel assessment of transfer reliability is conducted using absolute value percent transfer error, which is a standard measure of transfer error (Rosenberger and Stanley 2006). Transfer error *TE* is calculated as

$$(13) \quad TE = \frac{|\bar{V}_S - \bar{V}_T|}{\bar{V}_T} \cdot 100,$$

where $\bar{V}_S$ is a estimated mean WTP value from the assumed study site *S* and $\bar{V}_T$ is the estimated mean WTP value for the assumed policy or transfer site *T*, with the latter treated as the "true" value.[20] Again, we compare reliability for transfers conducted over goods measured in cardinal and relative terms.

As described above, we also verify whether our key findings are robust to the inclusion of socio-demographic controls in the model. Specifically, we re-estimate the model outlined with a set of added socio-demographic characteristics as explanatory variables influencing the means of random WTP parameters. Details and results for this model are provided in Appendix A. This supplementary model is provided for illustrative purposes, to demonstrate that the presented

---

[20] This calculation of *TE* is a form of convergent validity testing, comparing primary study and benefit transfer estimates as a measure of generalization error. This is typically done under the unstated assumption that a primary study provides a valid and reliable estimate of the true underlying value. However, primary studies may also be subject to measurement errors, and in some cases (e.g., very poor study design), may lead to *less* accurate estimates than a high quality benefit transfer. These and other issues related to generalization and measurement errors in benefit transfer are discussed by Rosenberger and Stanley (2006) and Rosenberger and Johnston (2009).

results are robust to the incorporation of socio-demographic variables.

Results

Table 2 presents model estimation results. *Status quo* is a binary indicator variable equal to one for the status quo (no program) option and zero otherwise. Similarly, the variables indicating emphasis on *Hard* (engineered) or *Soft* (natural) defenses are binary indicators and take a value of one when a variant emphasizes a given type of defense and zero otherwise. The remaining variables are continuous, as explained in table 1. As noted above, all continuous non-monetary attributes are defined in relative units for model estimation, so that one unit is equivalent to one percentage point. To ensure model convergence, continuous variables are scaled, with each variable representing a non-monetary attribute divided by 10 and *Cost* divided by 100.[21]

Before turning to benefit transfer, we briefly review the performance of the underlying WTP-space mixed logit model. The model is significant at $p<0.01$, with a McFadden pseudo-R^2 of 0.23. Most main effect coefficients are significant at $p<0.01$, and all estimated standard deviations are significant. Significant standard deviations suggest preference heterogeneity across households, justifying the use of the mixed logit model. The signs of estimated coefficients match prior expectations. We give primary emphasis to results for continuous variables in the model, as these are the variables that may be measured in either relative or cardinal units. Because attribute levels for *Homes, Wetlands* and *Beaches* are defined as losses

[21] This arbitrary scaling has no impact on model results, but enables the model to converge. When presenting WTP estimates subsequently (as in table 3), these estimates are simply rescaled by a factor of 10 for continuous variables and by a factor of 100 for discrete variables to reflect the original, native units of the model attributes. Attribute rescaling of this type is common when estimating WTP-space mixed logit models.

(table 1), the negative WTP estimates are expected. By reversing these negative signs, the resulting positive estimates may be interpreted as the WTP to prevent losses in these attributes.

As we control for differences in WTP between Old Saybrook and Waterford by interacting all preference parameters with a binary variable for Waterford, main effects reported in table 1 represent (scaled) WTP values for Old Saybrook respondents. Results for Old Saybrook indicate positive WTP to prevent the flooding of *Homes* and the loss of *Beaches* and *Wetlands*. However, we do not find statistically significant WTP in Old Saybrook for changes in *Seawall* miles, *ceteris paribus*. The next column of coefficient estimates shows the difference in (scaled) WTP estimates between Waterford and Old Saybrook. Results suggest statistically significant differences ($p<0.01$) between WTP estimates for the two communities for all attributes except *Beaches*. These findings have direct implications for benefit transfer, as explored below. By adding values from columns with main effects and Waterford interactions, one obtains (scaled) WTP estimates for Waterford respondents. Focusing on continuous attributes, WTP values for Waterford are larger compared to those for Old Saybrook for all attributes except for *Homes*. The positive coefficient on the interaction of the Waterford dummy with *Homes* suggests that, on average, Waterford respondents have lower WTP for the prevention of homes expected to flood than Old Saybrook respondents, though still they assign a positive value to this prevention.

To ease interpretation of the coefficients from table 2, table 3 reports re-calculated estimates of the marginal WTP values in dollars per unit change in each non-monetary attribute (i.e., reversing the variable scaling as described in footnote 21)—leading to WTP estimates with traditional interpretations. Table 3 also illustrates the results of traditional benefit-transfer validity tests, by testing the statistical equivalence of parallel, marginal welfare estimates across the two communities (Rosenberger 2015). Although not the primary focus of this article, we reject the null hypothesis of transfer validity (i.e., equal implicit prices) at $p<0.01$ for *Status quo* and *Soft* and at $p<0.05$ for *Hard*, defined as discrete (dummy) variables. These results are shown in the top panel of table 3. Because these attributes are always measured as binary variables, the associated tests do not provide insight into the effects of relative versus cardinal scaling. Hence, we present these results as background only.[22]

We now turn to the primary goal of the empirical assessment—to evaluate the effect of cardinal versus relative measurement of continuous attributes on transfer accuracy (validity and reliability). We first consider validity. As shown by table 1, each non-monetary continuous attribute was measured in both relative and cardinal units. Consequently, transfer validity can be evaluated in two ways, based on (a) marginal WTP measured per relative unit and (b) marginal WTP measured per cardinal unit. WTP estimates and hypothesis test results for this comparative assessment are provided in the lower panel of table 3. WTP estimates per relative unit are obtained directly from the model results reported in table 2, after accounting for the scaling needed for convergence. The WTP values per cardinal unit are calculated using equation (8) with

[22] The WTP estimates for the alternative specific constant (ASC) *Status quo* (table 3) are interpreted as fixed WTP associated with maintaining the status quo (here, negative), assuming no change in attributes from status quo levels. In some cases, estimated ASC parameters such as these reflect legitimate sources of utility not otherwise captured by choice attributes that are appropriate for inclusion in welfare analysis (Colombo and Hanley 2008). In other instances, these parameters may capture the effects of yea-saying or other effects that one might wish to exclude from benefit transfers even though ASCs are included in model estimation (Morrison et al. 2002).

information on attribute baselines from table 1.[23] Evaluation of associated hypotheses of equal WTP estimates across sites is based on Wald tests.

As shown by the lower panel of table 3, we reject the validity of benefit transfers between Old Saybrook and Waterford (i.e., reject the null hypothesis of equal implicit prices) for all continuous attributes measured in cardinal units ($p<0.01$). That is, when attributes are measured solely in cardinal units, there are no statistically valid transfers of implicit prices between Old Saybrook and Waterford. In contrast, the transfer of marginal WTP for attributes measured in relative units rejects the null hypothesis for three out of the four cases. We fail to reject the null hypothesis of equal implicit prices for *Beaches*, indicating validity of this benefit transfer between Old Saybrook and Waterford. That is, we find one statistically valid *relative* benefit transfer, but no valid *cardinal* transfers.

To complement the assessment of benefit transfer validity, we also compare the reliability of cardinal and relative transfers. This requires a comparison of transfer error magnitudes, with errors quantified and interpreted as in equation (13). Unlike validity tests—which simply verify the equivalence of two welfare measures—the outcome of reliability tests depends on the direction of the transfer (which determines the denominator of (13)). Hence, table 4 summarizes the errors calculated for transfers in both relative and cardinal units and for the two directions of transfers: from Old Saybrook to Waterford and from Waterford to Old Saybrook. To facilitate the comparison, table 4 also includes ratios of errors from cardinal transfers to corresponding errors from relative transfers, providing a consistent measure of the difference

[23] For example, a per household WTP estimate of -\$9.048 per 1% of homes flooded in Old Saybrook is equivalent to -\$9.048·(100/5,034) = -\$0.180 per home flooded. One can also conduct the same analysis using cardinal units and make a transformation to relative units, with equivalent results.

between errors associated with these two measurement conventions. Ratios greater than 1 imply that cardinal transfer errors are greater than relative transfer errors, and vice versa.

Reliability results parallel those on validity above, indicating greater reliability of benefit transfer when attributes are measured in relative units. In all cases tested, transfer errors are smaller (and in some, substantially smaller) for relative than for cardinal transfers. Errors from cardinal transfers are between 1.02 and 16.63 times greater than parallel errors from relative transfers. The ratio of cardinal transfer errors to relative transfer errors averages 1.87 for transfers from Old Saybrook to Waterford, and 5.36 for transfers from Waterford to Old Saybrook. That is, transfer errors are approximately *two- to five-times larger*, on average, when transfers are conducted over goods measured in cardinal units than in relative units.[24] Error differences of this magnitude exceed error differences associated with aspects of transfer methodology given much greater emphasis in the literature, such as the use of unit value versus benefit function transfer (Rosenberger 2015).

As an illustration, consider a transfer of implicit prices for *Wetlands* from Waterford to Old Saybrook. Transfers in relative units yield an average absolute value error of 53.61%. The same transfer in cardinal units yields a parallel error of 891.48%. This result is intuitive. Old Saybrook has a much larger number of current wetland acres than Waterford (497 versus 77; table 1), such that the implicit price for protecting each cardinal acre is lower ($1.67versus $16.55; table 3). Hence, a benefit transfer for *Wetlands* measured in cardinal acres generates large errors, whereas a transfer over relative units (percent of current wetland acres) generates smaller errors. These findings are consistent with implications of the theoretical model.

---

[24] Even if we exclude the largest ratio of 16.63 as an outlier (the transfer of WTP for *Wetlands* from Waterford to Old Saybrook), transfer errors remain between 60% and 87% larger for cardinal than for relative transfers.

These results are robust to different model specifications, including those that adjust for socio-demographic differences between the communities (see Appendix A) and those based on independent model estimation for each community (see Appendix B). In general, results from different specifications suggest that transfers over attributes measured in relative units outperform otherwise identical transfers over attributes measured in cardinal units. These empirical results are consistent with the theoretical model, which suggests that the typical inability of cardinal transfers to accommodate diminishing marginal utility (at least when estimated using linear utility functions of the type ubiquitous in choice experiments) can lead to large transfer errors when there are non-marginal differences in attribute baselines between sites. The same findings hold for unit value transfers as well as benefit function transfers that adjust for socio-demographic differences between communities.

These results have direct implications for the implementation and testing of environmental benefit transfer. First, if results such as these were shown to hold more broadly, it would imply that environmental goods should often be measured in relative terms for purposes of benefit transfers, *ceteris paribus*. This would not be a universal recommendation, but—following the theoretical model—would apply to a wide range of benefit transfer situations. Note that this approach would not generally require new or different data collection within primary studies—in many cases WTP measures for cardinally-measured goods can simply be rescaled during welfare estimation (as we have done here).[25] Second, these results imply that the apparent

[25] This requires that primary valuation studies report information on cardinal baselines. At least for stated preference studies, guidelines of Johnston et al. (2017) indicate that baselines should be presented clearly to respondents in questionnaires. If researchers follow this guideline, information on these baselines can subsequently be made available in published documents and/or provided directly to benefit transfer analysts upon request.

"transferability" of welfare estimates between sites—measured in either validity or reliability terms—can depend critically on whether goods are measured in cardinal or relative units. Consequently, past convergent validity tests in the literature that find invalid or large-error transfers for cardinally-measured goods might show *opposite (or at least different) results if goods were scaled differently prior to the test*. The potential sensitivity of results in the published literature to seemingly trivial variable rescaling suggests, at a minimum, that benefit transfer evaluations should consider this issue moving forward. It also implies that future summaries of benefit transfer validity and reliability (c.f., Rosenberger and Stanley 2006; Rosenberger 2015) could be enhanced by reporting variable measurement conventions.

Conclusion

Measurement conventions for environmental goods—such as whether goods are measured in cardinal or relative units—are often considered second-order or even trivial issues within non-market valuation. Only a few articles have studied ramifications of cardinal versus relative measurement for environmental valuation, and most of these address implications for scope effects in stated preference research. Here, we demonstrate that measurement conventions such as these can be of critical importance for benefit transfer. To our knowledge, this is the first acknowledgement of this type in the literature. Both the theoretical model and empirical analysis suggest that transfers over goods measured in relative units can often outperform transfers over otherwise identical goods measured in cardinal units. The results of benefit transfer convergent validity tests may also hinge on whether goods are measured in cardinal or relative units.

Although the empirical magnitude of these effects will vary across valuation contexts, the conceptual insights from the analysis apply to any non-market commodity (i.e., quantity or quality change) for which relevant commodity baselines vary across sites. Many types of quantity and quality changes considered by the non-market valuation literature fall into this category. There are some cases, however, where commodity measurement conventions would be less (or not) germane for benefit transfers. For example, commodities valued solely based on scope realized at a global scale[26] would have the same relevant baseline everywhere—such that transformations between cardinal and relative units would be immaterial for benefit transfer. Other types of commodities have values that are assumed by convention to be fixed for purposes of benefit-cost analysis within particular jurisdictions, with transfers always conducted in the same units. An example is the default central value of statistical life (VSL) estimate recommended for all US EPA benefit-cost analyses (US EPA 2014). Moreover, for any type of non-market commodity, the empirical impact of measurement conventions on transfer errors will vary depending on the type of benefit transfer conducted and the structure of estimated preference functions. Evaluations of variations of this type are left for future work.

Variations of this type notwithstanding, the presented results could have multiple implications for the implementation and evaluation of non-market valuation. For example, in addition to the direct repercussions for benefit transfer, results such as these might help valuation practitioners design studies that are better able to support subsequent benefit transfers—for instance, by quantifying goods in ways that enhance transfer prospects. In doing so, however, questions and seeming conflicts may arise. For example, as noted above, meta-analytic results of Ojea and Loureiro (2011) seem to imply the superiority of cardinally-measured goods for stated preference valuation, which could possibly lead to native welfare estimates that are less-well

[26] An example would be an endangered species that is valued solely for its existence at a global scale, such that only aggregate global populations are relevant to value.

suited to benefit transfer (when considered without rescaling into relative terms).[27]

Regardless, given the results presented above, there are some relatively simple steps that might be taken by researchers to improve the usefulness of their results for benefit transfers. For example, where feasible, researchers could present parallel sets of parameter estimates for changes measured in both cardinal and relative units, or provide the data necessary for analysts to make necessary transformations. Researchers could also make data and code available so that benefit transfer practitioners could re-estimate models with commodity changes rescaled as desired. Information of this type could be posted in supplemental online appendices or data repositories. Benefit transfer researchers, in turn, could assess the sensitivity of results to the use of cardinal versus relative transfers (where this is possible), and describe why a particular type of commodity measurement convention was used.

Results presented here must be viewed within the context of the present case study, along with implied assumptions and possible limitations. As noted above, additional research will be required to assess the extent to which similar findings apply to other study contexts, involving valuation of different types of goods over different populations. Subsequent research could also examine whether results are sensitive to the native measures used within the primary valuation study, that is, those used to communicate changes within original stated preference scenarios. In our survey, scenarios communicated changes in both cardinal and relative terms. Another limitation is that our study does not consider a case in which the transfer function itself allows direct adjustment of WTP for changes in the status quo (e.g., the current amount of marsh or beach in each community). Functions of this type are available through some primary studies and meta-analyses. In concept, such a function could at least potentially attenuate some of the

[27] Ojea and Loureiro (2011) do not account for varying baseline levels within their meta-analysis.

differences between cardinal and relative transfers. Issues such as these can be accommodated through straightforward extensions of the presented approach.

These and other caveats aside, the results presented above provide clear evidence that often overlooked, and deceivingly simple, choices such as measurement conventions for non-market goods can have first-order impacts on the validity and reliability of benefit transfer. Indeed, our empirical results suggest that measurement conventions can be among the most important determinants of transfer accuracy. Results such as these demonstrate that fundamental aspects of the transfer, including the basic methods used to quantify goods, can be more important for transfer accuracy than sophisticated aspects of transfer methodology. Ironically, the latter issues are given much greater attention in the literature.[28] Our results further suggest another dimension that researchers may wish to consider when designing primary valuation studies. Up to this point, guidance for commodity measurement in valuation has focused almost entirely on consequences for primary valuation studies (see citations above). Our results imply that these conventions may have a distinct and important set of consequences for benefit transfer.

[28] See, for example, the 2018 special issue "Benefit Transfer: Current Practice and Future Prospects," *Environmental and Resource Economics*, 69(3).

References

Babyak, Michael A. 2009. "Rescaling continuous predictors in regression models. Statistical Tips from the Editors of *Psychosomatic Medicine*." http://stattips.blogspot.com/2009/08/rescaling-continuous-predictors-in.html.

Barton, David N. 2002. "The transferability of benefit transfer: Contingent valuation of water quality improvements in Costa Rica." *Ecological Economics* 42(1-2): 147-164.

Bateman, Ian J., Roy Brouwer, Silvia Ferrini, Marije Schaafsma, David N. Barton, Alex Dubgaard, Berit Hasler, Stephanie Hime, Inge Liekens, Ståle Navrud, Leo De Nocker, Rasa Ščeponavičiūtė, and Daiva Semėnienė. 2011. "Making benefit transfers work: deriving and testing principles for value transfers for similar and dissimilar sites using a case study of the non-market benefits of water quality improvements across Europe." *Environmental and Resource Economics* 50(3): 365-387.

Bergstrom, John C., and Paul De Civita. 1999. "Status of benefit transfer in the United States and Canada: A review." *Canadian Journal of Agricultural Economics* 47(1): 79-87.

Bishop, Richard C. and Kevin J. Boyle. 2019. "Reliability and validity in nonmarket valuation." Environmental and Resource Economics 72(2): 559-582.

Boyd, James, Paul Ringold, Alan Krupnick, Robert J. Johnston, Matthew A. Weber, and Kim Hall. 2016. "Ecosystem services indicators: Improving the linkage between biophysical and economic analyses." *International Review of Environmental and Resource Economics* 8(3-4): 359-443.

Boyle, Kevin J., William H. Desvousges, F. Reed Johnson, Richard W. Dunford, and Sarah P. Hudson. 1994. "An investigation of part-whole biases in contingent-valuation studies." *Journal of Environmental Economics and Management* 27(1): 64-83.

Brouwer, Roy, Markus Bliem, Michael Getzner, Sandor Kerekes, Simon Milton, Teodora Palarie, Zsuzsanna Szerényi, Angheluta Vadineanu, and Alfred Wagtendonk. 2016. "Valuation and transferability of the non-market benefits of river restoration in the Danube river basin using a choice experiment." *Ecological Engineering* 87: 20-29.

Brouwer, Roy, and Ståle Navrud. 2015. "The use and development of benefit transfer in Europe." In: *Benefit Transfer of Environmental and Resource Values: A Guide for Researchers and Practitioners*. Ed. by Robert J. Johnston, John Rolfe, Randall S. Rosenberger, and Roy Brouwer. Dordrecht, The Netherlands: Springer.

Brouwer, Roy, and Frank A. Spaninks. 1999. "The validity of environmental benefits transfer: Further empirical testing." *Environmental and Resource Economics* 14(1): 95-117.

Carson, Richard T. 2012. "Contingent valuation: A practical alternative when prices aren't available." *Journal of Economic Perspectives* 26(4): 27-42.

Colombo, Sergio, and Nick Hanley. 2008. "How can we reduce the errors from benefits transfer? An investigation using the choice experiment method." *Land Economics* 84(1): 128-147.

Colombo, Sergio, Javier Calatrava-Requena, and Nick Hanley. 2007. "Testing choice experiment for benefit transfer with preference heterogeneity." *American Journal of Agricultural Economics* 89(1): 135-51.

Czajkowski, Mikołaj, Heini Ahtiainen, Janne Artell, and Jürgen Meyerhoff. 2017a. "Choosing a functional form for an international benefit transfer: Evidence from a nine-country valuation experiment." *Ecological Economics* 134: 104-113.

Czajkowski, Mikołaj, and Milan Ščasný. 2010. "Study on benefit transfer in an international setting. How to improve welfare estimates in the case of the countries' income heterogeneity?" *Ecological Economics* 69(12): 2409-2416.

Czajkowski, Mikołaj, Christian A. Vossler, Wiktor Budziński, Aleksandra Wiśniewska, and Ewa Zawojska. 2017b. “Addressing empirical challenges related to the incentive compatibility of stated preferences methods.” *Journal of Economic Behavior and Organization* 142: 47-63.

Daly, Andrew, Stephane Hess, and Kenneth Train. 2012. “Assuring finite moments for willingness to pay in random coefficient models.” *Transportation* 39(1): 19-31.

Desvousges, William H., F. Reed Johnson, Richard W. Dunford, Kevin J. Boyle, Sara P. Hudson, and Nicole Wilson. 1993. “Measuring natural resource damages with contingent valuation: Tests of validity and reliability.” In: *Contingent Valuation: A Critical Assessment*. Ed. by Jerry A. Hausman. Amsterdam, The Netherlands: North Holland.

Dillman, Don A., Jolene D. Smyth, and Leah M. Christian. 2009. *Internet, Mail, and Mixed-mode Surveys: The Tailored Design Method*. Toronto, Canada: Wiley Publishing.

Ferrini, Silvia, and Riccardo Scarpa. 2007. “Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study.” *Journal of Environmental Economics and Management* 53(3): 342-363.

Glenk, Klaus, Julia Martin-Ortega, Manuel Pulido-Velazquez, and Jacqueline Potts. 2015. “Inferring attribute non-attendance from discrete choice experiments: Implications for benefit transfer.” *Environmental and Resource Economics* 60(4): 497-520.

Griffiths, Charles, and William J. Wheeler. 2005. “Benefit-cost analysis of regulations affecting surface water quality in the United States.” In: *Cost-Benefit Analysis and Water Resources Management*. Ed. by Roy Brouwer and David Pearce. Cheltenham, UK: Edward Elgar.

Gujarati, Damodar, N. 1988. *Basic Econometrics*, 2nd Edition. New York, NY: McGraw-Hill.

Hanemann, W. Michael. 1984. “Welfare evaluation in contingent valuation experiments with discrete responses.” *American Journal of Agricultural Economics* 66(3): 332-341.

Hanemann, W. Michael. 1994. “Valuing the environment through contingent valuation.” *Journal of Economic Perspectives* 8(4): 19-43.

Hanley, Nick, Sergio Colombo, Dugald Tinch, Andrew Black, and Ashar Aftab. 2006. “Estimating the benefits of water quality improvements under the Water Framework Directive: Are benefits transferable?” *European Review of Agricultural Economics* 33(3): 391-413.

Holmes, Thomas P., Wiktor Adamowicz, and Fredrik Carlsson. 2017. Choice Experiments. In: *A Primer on Nonmarket Valuation*, 2nd Edition. Ed. by Patricia A. Champ, Kevin J. Boyle, and Thomas C. Brown. Dordrecht, The Netherlands: Springer.

Hoyos, David. 2010. “The state of the art of environmental valuation with discrete choice experiments.” *Ecological Economics* 69(8): 1595-1603.

Iovanna, Richard, and Charles Griffiths. 2006. “Clean water, ecological benefits, and benefits transfer: A work in progress at the U.S. EPA.” *Ecological Economics* 60(2): 473-482.

Johnston, Robert J., and Abdulallah S. Abdulrahman. 2017. “Systematic non-response in discrete choice experiments: Implications for the valuation of climate risk reductions.” *Journal of Environmental Economics and Policy* 6(3): 246-267.

Johnston, Robert J., Elena Y. Besedin, and Benedict M. Holland. 2019. “Modeling distance decay within valuation meta-analysis.” *Environmental and Resource Economics* 72(3): 657-690.

Johnston, Robert J., Elena Y. Besedin, and Ryan Stapler. 2017. "Enhanced geospatial validity for meta-analysis and environmental benefit transfer: An application to water quality improvements." *Environmental and Resource Economics* 68(2): 343-375.

Johnston, Robert J., Kevin J. Boyle, Wiktor (Vic) Adamowicz, Jeff Bennett, Roy Brouwer, Trudy Ann Cameron, W. Michael Hanemann, Nick Hanley, Mandy Ryan, Riccardo Scarpa, Roger Tourangeau, and Christian A. Vossler. 2017. "Contemporary guidance for stated preference studies." *Journal of the Association of Environmental and Resource Economists* 4(2): 319-405.

Johnston, Robert J., and Joshua M. Duke. 2008. "Benefit transfer equivalence tests with non-normal distributions." *Environmental and Resource Economics* 41(1): 1-23.

Johnston, Robert J. and Joshua M. Duke. 2009. "Willingness to pay for land preservation across states and jurisdictional scale: Implications for benefit transfer." *Land Economics* 85(2): 217–237.

Johnston, Robert J., and Joshua M. Duke. 2010. "Socioeconomic adjustments and choice experiment benefit function transfer: Evaluating the common wisdom." *Resource and Energy Economics* 32(3): 421-438.

Johnston, R.J., C. Makriyannis and A.W. Whelchel. 2018. "Using ecosystem service values to evaluate tradeoffs in coastal hazard adaptation." *Coastal Management* 46(4): 259-277.

Johnston, Robert J., and Randall S. Rosenberger. 2010. "Methods, trends and controversies in contemporary benefit transfer." *Journal of Economic Surveys* 24(3): 479-510.

Johnston, Robert J., John Rolfe, Randall S. Rosenberger, and Roy Brouwer. 2015. "Introduction to benefit transfer methods." In: *Benefit Transfer of Environmental and Resource Values:*

A Guide for Researchers and Practitioners. Ed. by Robert J. Johnston, John Rolfe, Randall S. Rosenberger, and Roy Brouwer. Dordrecht, The Netherlands: Springer.

Johnston, Robert J., John Rolfe and Ewa Zawojska. 2018. Benefit transfer of environmental and resource values: Progress, prospects and challenges. *International Review of Environmental and Resource Economics* 12(2-3): 177-266.

Johnston, Robert J., Eric T. Schultz, Kathleen Segerson, Elena Y. Besedin, and Mahesh Ramachandran. 2012. "Enhancing the content validity of stated preference valuation: The structure and function of ecological indicators." *Land Economics* 88(1): 102-120.

Kaplowitz, Michael D., Frank Lupi, and John P. Hoehn. 2004. "Multiple methods for developing and evaluating a stated-choice questionnaire to value wetlands." In: *Methods for Testing and Evaluating Survey Questionnaires*. Ed. by Stanley Presser, Jennifer M. Rothgeb, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer. Hoboken, NJ: Wiley.

King, Gary. 1986. "How not to lie with statistics: Avoiding common mistakes in quantitative political science." *American Journal of Political Science* 30(3): 666-687.

Kling, Catherine L., and Daniel J. Phaneuf. 2018. "How are scope and adding up relevant for benefits transfer?" *Environmental and Resource Economics* 69(3): 483-502.

Loomis, John B. 2015. "The use of benefit transfer in the United States." In: *Benefit Transfer of Environmental and Resource Values: A Guide for Researchers and Practitioners*. Ed. by Robert J. Johnston, John Rolfe, Randall S. Rosenberger, and Roy Brouwer. Dordrecht, The Netherlands: Springer.

Luisetti, Tiziana, Ian J. Bateman, and R. Kerry Turner. 2011. "Testing the fundamental assumption of choice experiments: Are values absolute or relative?" *Land Economics* 87(2): 284-296.

Makriyannis, Christos, Robert J. Johnston, and Adam W. Whelchel. 2018. "Are choice experiment treatments of outcome uncertainty sufficient? An application to climate risk reductions." *Agricultural and Resource Economics Review* 47(3): 419-451.

McFadden, Daniel. 1974. "Conditional logit analysis of qualitative choice behaviour." In: *Frontiers in Econometrics*. Ed. by Paul Zarembka. New York, NY: Academic Press.

Morrison, Mark, and Jeff Bennett. 2004. "Valuing New South Wales rivers for use in benefit transfer." *Australian Journal of Agricultural and Resource Economics* 48(4): 591-611.

Morrison, Mark, and Olvar Bergland. 2006. "Prospects for the use of choice modelling for benefit transfer." *Ecological Economics* 60(2): 420-428.

Morrison, Mark, Jeff Bennett, Russell Blamey, and Jordan Louviere. 2002. "Choice modeling and tests of benefit transfer." *American Journal of Agricultural Economics* 84(1): 161-170.

Newbold, Stephen, R. David Simpson, D. Matthew Massey, Matthew T. Heberling, William Wheeler, Joel Corona, and Julie Hewitt. 2018a. "Benefit transfer challenges: Perspectives from U.S. Practitioners." *Environmental and Resource Economics* 69(3): 467-482.

Newbold, Stephen C., Patrick J. Walsh, D. Matthew Massey, and Julie Hewitt. 2018b. "Using structural restrictions to achieve theoretical consistency in benefit transfers." *Environmental and Resource Economics* 69(3): 529-553.

Ojea, Elena, and Maria L. Loureiro. 2011. "Identifying the scope effect on a meta-analysis of biodiversity valuation studies." *Resource and Energy Economics* 33(3): 706-724.

Proctor, Bernadette D., Jessica L. Semega, and Melissa A. Kollar. 2016. U.S. Census Bureau, Current Population Reports, P60-256(RV), Income and Poverty in the United States: 2015, U.S. Government Printing Office, Washington, DC.

Rosenberger, Randall S. 2015. "Benefit transfer validity and reliability." In: *Benefit Transfer of Environmental and Resource Values: A Guide for Researchers and Practitioners*. Ed. by Robert J. Johnston, John Rolfe, Randall S. Rosenberger, and Roy Brouwer. Dordrecht, The Netherlands: Springer.

Rosenberger, Randall S., and John B. Loomis. 2003. "Benefit transfer." In: *A Primer on Nonmarket Valuation*. Ed. by Patricia A. Champ, Kevin J. Boyle, and Thomas C. Brown. Dordrecht, The Netherlands: Kluwer.

Rosenberger, Randall S., and Tom D. Stanley. 2006. "Measurement, generalization and publication: Sources of error in benefit transfers and their management." *Ecological Economics* 60(2): 372-378.

Sándor, Zsolt, and Michel Wedel. 2001. "Designing conjoint choice experiments using managers' prior beliefs." *Journal of Marketing Research* 38(4): 430-444.

Sándor, Zsolt, and Michel Wedel. 2002. "Profile construction in experimental choice designs for mixed logit models." *Marketing Science* 21(4): 455-475.

Scarpa, Riccardo, and John M. Rose. 2008. "Design efficiency for non-market valuation with choice modelling: How to measure it, what to report and why." *Australian Journal of Agricultural and Resource Economics* 52(3): 253-282.

Scarpa, Riccardo, Mara Thiene, and Kenneth Train. 2008. "Utility in willingness to pay space: A tool to address confounding random scale effects in destination choice to the Alps." *American Journal of Agricultural Economics* 90(4): 994-1010.

Schkade, David A., and John W. Payne. 1994. "How people respond to contingent valuation questions: A verbal protocol analysis of willingness to pay for an environmental regulation." *Journal of Environmental Economics and Management* 26(1): 88-109.

Smith, V. Kerry, George Van Houtven, and Subhrendu K. Pattanayak. 2002. "Benefit transfer via preference calibration: 'Prudential algebra' for policy." *Land Economics* 78(1): 132-152.

Taylor, Laura O., 2017. "Hedonics." In: A Primer on Nonmarket Valuation, 2nd Edition. Ed by by Patricia A. Champ, Kevin J. Boyle, and Thomas C. Brown. Dordrecht, The Netherlands: Springer.

Thiene, Mara, and Riccardo Scarpa. 2009. "Deriving and testing efficient estimates of WTP distributions in destination choice models." *Environmental and Resource Economics* 44(3): 379-395.

Train, Kenneth. 2009. *Discrete Choice Methods with Simulation*. New York, NY: Cambridge University Press.

Train, Kenneth, and Melvyn Weeks. 2005. "Discrete choice models in preference space and willing-to-pay space." In: *Applications of Simulation Methods in Environmental and Resource Economics*. Ed. by Riccardo Scarpa and Anna Alberini. Dordrecht, the Netherlands: Springer.

US Environmental Protection Agency (US EPA). 2011. Economic and Benefits Analysis for Proposed Section 316(b) Existing Facilities Rule. EPA 821-R-11-002. March 28. Washington, DC.

US Environmental Protection Agency (US EPA). 2014. Guidelines for Preparing Economic Analysis. National Center for Environmental Economics, Office of Policy, Washington, DC.

US Environmental Protection Agency and US Department of the Army (US EPA and US DOA). 2015. Economic Analysis of the EPA-Army Clean Water Rule. May 20. Downloaded August 9, 2019 from https://www.epa.gov/sites/production/files/2015-06/documents/508-final_clean_water_rule_economic_analysis_5-20-15.pdf.

Table 1. Attributes in the DCEs

Attribute	Description	Attribute Levels in Old Saybrook	Attribute Levels in Waterford
Homes	Number of homes expected to flood in a storm of Category 3 or higher in the mid-2020s	Out of the total of 5,034 homes: 1,812 (36%) 2,165 (43%) 2,585 (51%) SQ 2,970 (59%)	Out of the total of 8,460 homes: 169 (2%) 338 (4%) 566 (7%) SQ 846 (10%)
Wetlands (Costal Marshes)	Acres of coastal marshes expected to be lost by the mid-2020s due to flooding or erosion	Out of the total of 497 acres: 10 (2%) 25 (5%) SQ 50 (10%)	Out of the total of 77 acres: 4 (5%) 9 (12%) SQ 15 (19%)
Beaches	Acres of beaches and dunes expected to be lost by the mid-2020s due to flooding or erosion	Out of the total of 30 acres: 1 (4%) 3 (10%) SQ 5 (16%)	Out of the total of 36 acres: 1 (4%) 4 (10%) SQ 6 (16%)
Seawalls	Miles of the coastline shielded by hard defenses by the mid-2020s	Out of the total of 50 miles: 8 (15%) 12 (24%) SQ 18 (35%)	Out of the total of 26 miles: 10 (40%) 13 (50%) SQ 16 (60%)
Emphasis of adaptation efforts	Type of defenses (hard or soft) with more emphasis placed in the	More emphasis on hard defenses	More emphasis on hard defenses

(*Hard*; *Soft*)	adaptation policy relative to the status quo	More emphasis on soft defenses No change in the emphasis [SQ]	More emphasis on soft defenses No change in the emphasis [SQ]
Cost	Household annual cost in a form of an unavoidable increase in taxes and fees required to implement the adaptation policy	\$0 [SQ] \$35 \$65 \$95 \$125 \$155	\$0 [SQ] \$35 \$65 \$95 \$125 \$155

Note: The levels associated with the status quo (no change) variant are marked with a superscript SQ.

Table 2. Mixed Logit Model in WTP-Space

	Mean: Main Effect	Mean: Shifter for Waterford	Standard Deviation
Status quo	-1.3371***	-0.3537***	3.3205***
	(0.0737)	(0.1092)	(0.2236)
Homes	-0.9048***	0.8664***	0.9503***
	(0.0613)	(0.2063)	(0.0618)
Wetlands	-0.8294***	-0.4447***	0.9284***
	(0.1389)	(0.1392)	(0.0799)
Beaches	-0.7664***	-0.0490	1.2332***
	(0.1323)	(0.1464)	(0.0879)
Seawalls	0.0154	0.1717*	0.6894***
	(0.0796)	(0.0917)	(0.0650)
Emphasis on *Hard* defenses	-0.4359***	0.3057**	1.0721***
	(0.1198)	(0.1544)	(0.0976)
Emphasis on *Soft* defenses	0.1569	-0.5596***	1.3502***
	(0.1348)	(0.2158)	(0.1030)
Cost	2.6980***	-1.2708*	2.3532***
	(0.8321)	(0.6571)	(0.4945)
Log-likelihood	-685.92		
Log-likelihood with constants only	-893.49		
McFadden pseudo-R^2	0.2323		
Ben-Akiva-Lerman pseudo-R^2	0.4549		

Number of observations	815

Note: Standard errors are given in brackets. ***, ** and * imply significance at the 1%, 5% and 10% level, respectively. The estimates for the monetary attribute *Cost* are for the underlying lognormal distribution. We use a negative of *Cost* in the estimation.

Table 3. Marginal WTP for Non-Monetary Attributes

Attributes defined as discrete variables

	WTP_{OS}	WTP_{W}	$WTP_{OS} \neq WTP_{W}$
Status quo	-133.706	-169.072	***
Emphasis on *Hard* defenses	-43.594	-13.025	**
Emphasis on *Soft* defenses	15.688	-40.273	***

Attributes defined as continuous variables

	WTP per Relative Unit (per one percentage point change)			WTP per Cardinal Unit (units indicated in square brackets)		
	WTP_{OS}	WTP_{W}	$WTP_{OS} \neq WTP_{W}$	WTP_{OS}	WTP_{W}	$WTP_{OS} \neq WTP_{W}$
Homes [number]	-9.048	-0.384	***	-0.180	-0.005	***
Wetlands [acre]	-8.294	-12.741	***	-1.669	-16.547	***
Beaches [acre]	-7.664	-8.154		-25.547	-22.650	***
Seawalls [mile]	0.154	1.871	*	0.308	7.195	***

Note: WTP_{OS} denotes WTP of Old Saybrook residents and WTP_{W} denotes WTP of Waterford residents. ***, ** and * imply that WTP_{OS} and WTP_{W} are statistically different at the 1%, 5% and 10% significance level, respectively. For WTP in relative units and for discrete-defined attributes, verification of statistical significance of the difference in WTP between the two sites is based on significance of the coefficients of the interaction terms in the model reported in table 2. For WTP in cardinal units, this verification is based on Wald tests.

Table 4. Benefit Transfer Errors (TE) for Attributes Measured on Continuous Scales

	Transfer from Old Saybrook to Waterford			Transfer from Waterford to Old Saybrook		
	TE for **relative** transfer	TE for **cardinal** transfer	TE ratio (cardinal / relative TE)	TE for **relative** transfer	TE for **cardinal** transfer	TE ratio (cardinal / relative TE)
Homes	2,258.17%	3,863.08%	1.71	95.76%	97.48%	1.02
Wetlands	34.90%	89.91%	2.58	53.61%	891.48%	16.63
Beaches	6.01%	12.79%	2.13	6.39%	11.34%	1.78
Seawalls	91.78%	95.72%	1.04	1,116.22%	2,238.88%	2.01
Mean TE ratio			1.87			5.36

PLEASE VOTE

Question 4.

PROTECTION OPTION A and **PROTECTION OPTION B** are possible protection options for Waterford. **NO NEW ACTION** shows what is expected to occur with no additional protection. All plans involve hard and soft defenses in different areas. Given a choice among the three, how would you vote?

Methods and Effects of Protection	Result in 2020s with NO NEW ACTION	Result in 2020s with PROTECTION OPTION A	Result in 2020s with PROTECTION OPTION B
	No Change in Existing Defenses	SIMILAR Emphasis on Hard and Soft Defenses	More Emphasis on SOFT Defenses
Homes Flooded	**7%** 566 of 8,460 homes expected to flood in a Category 3 storm	**7%** 566 of 8,460 homes expected to flood in a Category 3 storm	**7%** 566 of 8,460 homes expected to flood in a Category 3 storm
Wetlands Lost	**12%** 9 of 77 wetland acres expected to be lost	**12%** 9 of 77 wetland acres expected to be lost	**5%** 4 of 77 wetland acres expected to be lost
Beaches and Dunes Lost	**10%** 4 of 36 beach acres expected to be lost	**4%** 1 of 36 beach acres expected to be lost	**4%** 1 of 36 beach acres expected to be lost
Seawalls and Coastal Armoring	**50%** 13 of 26 miles of coast armored	**50%** 13 of 26 miles of coast armored	**40%** 10 of 26 miles of coast armored
$ Cost to Your Household per Year	**$0** Increase in annual taxes or fees	**$95** Increase in annual taxes or fees	**$125** Increase in annual taxes or fees
HOW WOULD YOU VOTE? (CHOOSE ONLY ONE) I vote for	☐ I vote for **NO NEW ACTION**	☐ I vote for **PROTECTION OPTION A**	☐ I vote for **PROTECTION OPTION B**

Figure 1. An example of a choice task (Waterford version)

Appendix A. Comparison Results Using Models with Socio-Demographic Interactions

The Old Saybrook and Waterford samples differ in socio-demographic characteristics (table A1). To evaluate whether these differences are salient to our conclusions, this appendix repeats the analysis above using a benefit function transfer that controls for a set of socio-demographic variables. The model enables adjustments for binary indicator variables identifying female respondents, respondents living in households with total annual income of at least $80,000, respondents aged 60 or more, and respondents who have obtained an academic degree. The cutoffs for income and age approximate median levels in the sample, and for these cutoffs, we observe statistically significant differences between Old Saybrook and Waterford samples. These socio-demographic variables were chosen to reflect the type of adjustments made commonly within benefit transfers (i.e., for gender, income, age and education; Colombo, Calatrava-Requena and Hanley 2007; Johnston and Duke 2010), and for which data are typically available to researchers without conducting primary research at policy sites. This supplementary analysis is presented for illustrative purposes to ascertain whether our main findings are robust.

The model is structurally parallel to the mixed logit model described in the main text, but allows all WTP parameters to vary systematically as a function of the four socio-demographic variables. This is accommodated using the same structure applied to the binary indicator variable for Waterford, W_h, in equations (11) and (12). Results are provided in table A2. As in the main text, we assess validity and reliability of transfers employing cardinal and relative measurement conventions. Following standard benefit function transfer protocols (Johnston et al. 2015), mean values of the socio-demographic characteristics specific to Old Saybrook and Waterford samples are used for benefit transfers to those communities, thereby adjusting transfers for the socio-demographic conditions at each site. Results are shown by tables A3 and A4.

Results parallel those presented in the main text—transfers in relative units outperform transfers in cardinal units. Transfer validity is rejected in all cases, for all attributes, except for *Beaches* in relative transfers (table A3). Average transfer errors in table A4 are smaller for relative transfers, with average transfer error ratios between 3.16 and 11.17 (table A4). All transfer errors are larger for cardinal transfers, with the exception of two cases for which transfer errors are nearly identical (transferring WTP for *Seawalls* from Old Saybrook to Waterford and WTP for *Homes* from Waterford to Old Saybrook). These results align with expectations—there is no reason to expect the inclusion of socio-demographic characteristics to influence the comparative performance of relative versus cardinal transfers.[29]

[29] Results in Table A2 imply that one small sub-group of respondents (Waterford male respondents aged 60 or more, with income below $80,000 and no academic degree) holds a positive and significant WTP for *increases* in the number of community-wide homes at risk of flooding. (For a few other sub-groups this WTP is positive but not significantly different from zero.) Only nine respondents in the sample are within this group, and homes of these respondents are at low risk of flooding (i.e., they have not experienced recent flood damage to their homes and are not located in flood zones). Although this finding might seem superficially counter-intuitive, it is consistent with preferences expressed in focus groups, as explained by Johnston, Makriyannis and Whelchel (2018). Specifically, residents such as these often view the protection of private homes as the *responsibility of individual homeowners*—not a valued outcome for which (not-at-risk) residents are willing to pay. These residents would prefer *not* to have tax revenues spent on the protection of other people’s (often waterfront) homes, leading to zero or positive WTP for this attribute.

.

Table A1. Basic Socio-Demographic Characteristics of Waterford and Old Saybrook Respondents

Variable	Definition	Old Saybrook	Waterford	P-value
Female	1 for females, 0 otherwise (including omitted responses)	44.36%	51.60%	0.000
Age	Respondent's age	60.22 (13.58)	55.87 (14.56)	0.000
Age ≥ 60	1 for respondents aged 60 or more, 0 otherwise (including omitted responses)	59.02%	37.25%	0.000
Academic degree	1 for respondents with some academic degree, 0 otherwise (including omitted responses)	72.06%	66.83%	0.005
Annual income	Total household annual income in USD	121,880 (74,866)	106,446 (64,067)	0.000
Income ≥ $80,000	1 for respondents living in households with total annual income of at least $80,000, 0 otherwise (including omitted responses)	52.94%	47.42%	0.006

Note: P-values are for the null hypothesis of no difference between Old Saybrook and Waterford samples with respect to a given socio-demographic characteristic. For *Female*, *Age ≥ 60*, *Academic degree*, and *Income ≥ $80,000*, the table shows the shares of participants within each sample. For these variables, chi-squared tests of equality of proportions are conducted. For *Age* and *Annual income*, means (and standard deviations in the brackets) are reported, and the Wilcoxon signed-rank test is used to evaluate significance of the difference. Data on income were collected through a discrete choice question including eight income categories. Here, the responses are translated into mid-points of the income intervals presented in the categories.

Table A2. Mixed Logit Model in WTP-Space

	Mean: Main Effect	**Mean: Shifter for Waterford**	**Mean: Interaction with *Female***	**Mean: Interaction with *Income ≥ $80,000***	**Mean: Interaction with *Age ≥ 60***	**Mean: Interaction with *Academic Degree***	**Standard Deviation**
Status quo	-0.1212	-0.3241***	-1.0952***	0.1101	-0.4564***	-0.6608***	3.2202***
	(0.1311)	(0.0582)	(0.0815)	(0.0833)	(0.0609)	(0.0813)	(0.1265)
Homes	-0.9769***	1.0194***	-0.3439***	-0.2390***	0.9963***	-0.2705***	0.8797***
	(0.0995)	(0.1081)	(0.0512)	(0.0664)	(0.0611)	(0.0765)	(0.0431)
Wetlands	-0.9344***	-0.1538**	-0.6364***	0.3070***	0.3583***	-0.1606	0.7891***
	(0.1890)	(0.0608)	(0.0913)	(0.1052)	(0.0781)	(0.1123)	(0.0440)
Beaches	-0.9082***	0.0798	0.0909	-0.1702**	0.5024***	-0.1601	0.8259***
	(0.1790)	(0.0789)	(0.0837)	(0.0817)	(0.0813)	(0.1101)	(0.0449)
Seawalls	-0.2393*	0.2195***	0.1827***	0.0796	0.1288*	0.0454	0.6542***
	(0.1293)	(0.0584)	(0.0536)	(0.0578)	(0.0669)	(0.0749)	(0.0358)
Emphasis on *Hard*	-0.7970***	0.3427***	-0.1642*	0.5721***	0.4800***	0.0336	1.2741***
defenses	(0.1987)	(0.0734)	(0.0917)	(0.1324)	(0.0834)	(0.1336)	(0.0676)

Emphasis on *Soft*	-0.0795	-0.4815***	0.2283**	0.8653***	0.3981***	-0.6779***	1.4004***
defenses	(0.2161)	(0.1431)	(0.1032)	(0.1663)	(0.0902)	(0.1422)	(0.0640)
Cost	0.9491	-0.6442	0.5741	1.3848*	1.4267**	0.3996	2.4499***
	(0.9582)	(0.6455)	(0.6263)	(0.7557)	(0.6836)	(0.5926)	(0.6340)
Log-likelihood	-658.97						
Log-likelihood with constants only	-893.49						
McFadden pseudo-R^2	0.2625						
Ben-Akiva-Lerman pseudo-R^2	0.4721						
Number of observations	815						

Note: Standard errors are given in brackets. ***, ** and * imply significance at the 1%, 5% and 10% level, respectively. The estimates for the monetary attribute *Cost* are for the underlying lognormal distribution. We use a negative of *Cost* in the estimation.

Table A3. Marginal Transferred and Actual WTP Values for Non-Monetary Attributes

	Benefit Function Transfer from Old Saybrook to Waterford					
	WTP measured per **relative** unit			WTP measured per **cardinal** unit		
	WTP$_{\text{TRANSFER}}$	**WTP$_{\text{TRUE}}$**	**WTP$_{\text{TRANSFER}}$ ≠ WTP$_{\text{TRUE}}$**	**WTP$_{\text{TRANSFER}}$**	**WTP$_{\text{TRUE}}$**	**WTP$_{\text{TRANSFER}}$ ≠ WTP$_{\text{TRUE}}$**
Status quo	-122.523	-154.936	***			
Homes	-11.219	-1.025	***	-0.223	-0.012	***
Wetlands	-10.337	-11.875	**	-2.080	-15.422	***
Beaches	-8.936	-8.139		-29.787	-22.607	***
Seawalls	-0.141	2.054	***	-0.282	7.901	***
Emphasis on *Hard* defenses	-30.239	4.029	***			
Emphasis on *Soft* defenses	30.549	-17.602	***			

	Benefit Function Transfer from Waterford to Old Saybrook					
	WTP measured per **relative** unit			WTP measured per **cardinal** unit		
	WTP$_{\text{TRANSFER}}$	**WTP$_{\text{TRUE}}$**	**WTP$_{\text{TRANSFER}}$ ≠ WTP$_{\text{TRUE}}$**	**WTP$_{\text{TRANSFER}}$**	**WTP$_{\text{TRUE}}$**	**WTP$_{\text{TRANSFER}}$ ≠ WTP$_{\text{TRUE}}$**
Status quo	-160.687	-128.274	***			
Homes	1.314	-8.880	***	0.016	-0.176	***
Wetlands	-10.798	-9.261	**	-14.024	-1.863	***
Beaches	-7.150	-7.947		-19.861	-26.491	***
Seawalls	2.205	0.010	***	8.482	0.021	***
Emphasis on	14.350	-19.918	***			

Hard defenses			
Emphasis on *Soft* defenses	-16.392	31.758	***

Note: $WTP_{TRANSFER}$ denotes the transferred WTP value to the policy site and WTP_{TRUE} denotes the true, actual WTP value at the policy site. *** and ** imply that $WTP_{TRANSFER}$ and WTP_{TRUE} are statistically different at the 1% or 5% significance level, respectively. The verification of statistical significance is based on Wald tests.

Table A4. Benefit Transfer Errors (TE) for Attributes Measured on Continuous Scales

	Transfer from Old Saybrook to Waterford			Transfer from Waterford to Old Saybrook		
	TE for **relative** transfer	TE for **cardinal** transfer	TE ratio (cardinal / relative TE)	TE for **relative** transfer	TE for **cardinal** transfer	TE ratio (cardinal / relative TE)
Homes	994.73%	1,739.77%	1.75	114.80%	108.80%	0.95
Wetlands	12.95%	86.51%	6.68	16.60%	652.63%	39.31
Beaches	9.80%	31.76%	3.24	10.03%	25.03%	2.49
Seawalls	106.85%	103.56%	0.97	21,393.92%	41,234.46%	1.93
Mean TE ratio			3.16			11.17

Appendix B. Comparison Results Using Independent Models for Each Community

The benefit transfer assessment conducted in the main body of the paper relies on WTP values from a mixed logit model estimated on pooled data for Old Saybrook and Waterford. The pooled-model approach enables statistical efficiency gains and direct tests of WTP estimates between the two communities. However, when conducting real-world benefit transfers, researchers typically do not have access to underlying valuation study data (for the study site), and no valuation data are available for the policy site. Therefore, estimation of pooled models is not possible. Reflecting this reality, as a complementary analysis and a robustness check, this appendix repeats the analysis in the main text using WTP estimated via two independent mixed logit models for each community.

The mixed logit models presented here (table B1) are specified in parallel manner to the primary model in main text, with the model for each community following equation (10). The models are estimated via simulated maximum likelihood with 1,000 Sobol draws in WTP-space with non-monetary parameters specified as normally distributed and the cost parameter assumed to follow a lognormal distribution. As in the main model, continuous variables are scaled, with each variable representing a non-monetary attribute divided by 10 and Cost divided by 100. This is done to ensure model convergence.

Results support the robustness of conclusions drawn in the main text (tables B2 and B3). As expected, non-trivial differences in econometric structure between the pooled and independent models lead to modest changes in WTP point estimates.[30] Despite these modest differences, the conclusions of the analysis are robust—relative transfers to outperform cardinal

[30] For example, the pooled model imposes the same parameter standard deviations on both communities, whereas the unpooled models allow these parameter standard deviations to differ across the two communities.

transfers. Relative transfers are more often valid—validity is rejected for 75% of cardinal transfers but only for 50% of relative transfers (table B2). Transfer errors are also smaller for relative transfers (table B3), again mirroring results in the main text.

Table B1. Mixed logit models in WTP-space

	Old Saybrook		**Waterford**	
	Mean: Main effect	**Standard deviation**	**Mean: Main effect**	**Standard deviation**
Status quo	-1.4743***	5.2615***	-1.5036***	2.5644***
	(0.0635)	(0.0518)	(0.1321)	(0.3690)
Homes	-1.4365***	1.6950***	-0.1726	0.8864***
	(0.0483)	(0.0614)	(0.2207)	(0.2862)
Wetlands	-1.1533***	2.7431***	-0.9217***	0.6923***
	(0.0586)	(0.2415)	(0.1684)	(0.1730)
Beaches	-1.1627***	1.5318***	-0.7387***	0.9111***
	(0.0643)	(0.0873)	(0.2089)	(0.1568)
Seawalls	-0.0292	0.8024***	-0.0273	0.5773***
	(0.0329)	(0.0871)	(0.1255)	(0.1328)
Emphasis on *Hard*	-0.9069***	2.2031***	0.1059	1.0117***
defenses	(0.1486)	(0.0722)	(0.1772)	(0.1469)
Emphasis on *Soft*	0.2382***	2.2082***	-0.3424	2.0898***
defenses	(0.0759)	(0.0768)	(0.2496)	(0.3746)
Cost	8.6754***	5.7289***	1.6988***	2.2281***

	(1.1162)	(1.0273)	(0.6437)	(0.5753)
Log-likelihood	-325.86		-336.46	
Log-likelihood with constants only	-447.07		-446.38	
McFadden pseudo-R^2	0.2711		0.2463	
Ben-Akiva-Lerman pseudo-R^2	0.4698		0.4621	
Number of observations	408		407	

Note: Standard errors are given in brackets. *** implies significance at the 1% level. The estimates for the monetary attribute *Cost* are for the underlying lognormal distribution.

Table B2. Marginal WTP for Non-Monetary Attributes

Attributes defined as discrete variables

	WTP_{OS}	**WTP_{W}**	**$WTP_{OS} \neq WTP_{W}$**
Status quo	-147.435	-150.364	
Emphasis on *Hard* defenses	-90.690	10.591	***
Emphasis on *Soft* defenses	23.823	-34.242	**

Attributes defined as continuous variables

	WTP per relative unit (per one percentage point change)			**WTP per cardinal unit** (units indicated in square brackets)		
	WTP_{OS}	**WTP_{W}**	**$WTP_{OS} \neq WTP_{W}$**	**WTP_{OS}**	**WTP_{W}**	**$WTP_{OS} \neq WTP_{W}$**
Homes [number]	-14.365	-1.726	***	-0.285	-0.020	***
Wetlands [acre]	-11.533	-9.217		-2.321	-11.970	***
Beaches [acre]	-11.627	-7.387	*	-38.756	-20.519	***
Seawalls [mile]	-0.292	-0.273		-0.584	-1.051	

Note: WTP_{OS} denotes WTP of Old Saybrook residents and WTP_{W} denotes WTP of Waterford residents. ***, ** and * imply that WTP_{OS} and WTP_{W} are statistically different at the 1%, 5% and 10% significance level, respectively. Verification of statistical significance is based on z-tests.

Table B3. Benefit Transfer Errors (TE) for Attributes Measured on Continuous Scales

	Transfer from Old Saybrook to Waterford			**Transfer from Waterford to Old Saybrook**		
	TE for **relative** transfer	TE for **cardinal** transfer	TE ratio (cardinal / relative TE)	TE for **relative** transfer	TE for **cardinal** transfer	TE ratio (cardinal / relative TE)
Homes	732.52%	1299.11%	1.77	87.99%	92.85%	1.06
Wetlands	25.13%	80.61%	3.21	20.08%	415.83%	20.71
Beaches	57.40%	88.87%	1.55	36.47%	47.05%	1.29
Seawalls	6.80%	44.46%	6.54	6.37%	80.06%	12.57
Mean TE ratio			3.27			8.91