



**RESOURCES**  
*for the* **FUTURE**

# **Between Two Worlds: Methodological and Subjective Differences in Climate Impact Meta-Analyses**

**Peter H. Howard and Thomas Sterner**

**WP 22-10**  
**June 2022**



## About the Authors

**Peter H. Howard** is the economics director at the Institute for Policy Integrity at New York University School of Law. Much of his work focuses on the social cost of carbon and integrated assessment models. He is the lead researcher for the Cost of Carbon Pollution Project, a collaboration with the Environmental Defense Fund (EDF) and the Natural Resource Defense Council (NRDC). He holds a Ph.D. in Agricultural and Resource Economics from the University of California, Davis, where his research focused on climate change, environmental policy, and agricultural policy. Howard also holds a Bachelor of Arts from Bard College.

**Thomas Sterner** is a university fellow at Resources for the Future (RFF) and a Professor of Environmental Economics at the University of Gothenburg in Sweden. His work is focused on the design of policy instruments to deal with resource and environmental problems. Sterner has published more than a dozen books and a hundred articles on environmental policy instruments with applications to energy, climate, industry, transport economics and resource management in developing countries. He has in particular worked on the importance of fuel taxation for climate and transport policies. He has also worked on the feasibility and income distributional aspects of environmental policies. Another body of his work is focused on intertemporal discounting.

## Acknowledgements

We are grateful for stimulating discussions with Bill Nordhaus that formed the genesis for this project. In preparation for his Nobel Lecture, Nordhaus asked us to explain the differences between Nordhaus and Moffat (2017) and Howard and Sterner (2017). The resulting extended dialogue formed the starting point for this article. We would also like to thank John Hassler for some thought-provoking discussion during the early stages of the paper. Finally, we would like to thank Dr. Suhail Doi for his help and suggestions on applying his quality effects estimator in the climate damage context. Thomas Sterner would like to thank Formas Chips and Mistra Carbon Exit for research funding.

# About RFF

Resources for the Future (RFF) is an independent, nonprofit research institution in Washington, DC. Its mission is to improve environmental, energy, and natural resource decisions through impartial economic research and policy engagement. RFF is committed to being the most widely trusted source of research insights and policy solutions leading to a healthy environment and a thriving economy.

Working papers are research materials circulated by their authors for purposes of information and discussion. They have not necessarily undergone formal peer review. The views expressed here are those of the individual authors and may differ from those of other RFF experts, its officers, or its directors.

# Sharing Our Work

Our work is available for sharing and adaptation under an Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license. You can copy and redistribute our material in any medium or format; you must give appropriate credit, provide a link to the license, and indicate if changes were made, and you may not apply additional restrictions. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. You may not use the material for commercial purposes. If you remix, transform, or build upon the material, you may not distribute the modified material. For more information, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

# Abstract

In his 2019 Nobel Prize acceptance paper, William Nordhaus (2019) highlighted the uncertainty over climate damages by using two completely different damage functions: Nordhaus and Moffat (2017) and Howard and Sterner (2017). Despite their vastly different implications for climate policies, both were estimated using the meta-analysis technique, a method long considered the objective and scientifically rigorous way for combining results from multiple studies to develop a consensus estimate. This paper demonstrates that this disparity stems from differences in both methodological decisions (addressing methodological impacts and heteroskedasticity) and subjective decisions (about data search, selection, and weighting). Combining the two data sets, applying Nordhaus's quality weights, and applying the best methodological practices, we find damages of approximately 7 to 10 percent of gross domestic product (GDP) for a 3°C increase in global average surface temperature, depending on the inclusion of catastrophic and productivity impacts; the result is relatively robust to alternative data selection, weighting, and methodological assumptions. However, subjective differences between the weighting assumptions of the two earlier studies are still unresolved. To address subjectivity, this paper makes transparent existing weighting rules, develops new weighting rules, and applies a recently developed quality effects estimator from the medical literature (Doi et al. 2015). Operationalizing these rules, we demonstrate that damages are approximately 7 to 16 percent of GDP for a 3°C increase, though the upper end of this range, which includes catastrophic and productivity impacts, is sensitive to the selected model and weight specification.

# Contents

1. Introduction	1
2. Differences between Nordhaus and Moffat (2017) and Howard and Sterner (2017)	2
2.1 Data: Search, Selection, and Weighting	3
2.2 Methodological Differences	7
3. Analyzing Regression Differences	10
3.1 Replication of Nordhaus and Moffat (2017)	10
3.2 One-by-One Regressions Analyzing HS	11
3.3 Outliers	12
3.4 Takeaways	14
4. Improving Meta-Analysis: Addressing Methodological and Data Discrepancies	15
5. Addressing Subjectivity in Weights	18
5.1 Developing Objective Components of Weights	18
5.2 Subjective Assumptions Underlying Weights	21
5.3 Quality Effects Estimator	22
5.4 Results	23
6. Conclusion	25
7. References	27
8. Tables and Figures	32

# 1. Introduction

In his 2019 Nobel Prize acceptance speech, William Nordhaus (2019) argues that predicting climate damages is “the most difficult task” associated with the study of global warming and suffers the greatest uncertainties. Contrasting the DICE-2016R2 damage function (Nordhaus and Moffat 2017) with the damage function estimated by the authors of this paper (Howard and Sterner 2017), he notes that summaries of global damage estimates are “all over the map,” with a range of 1.9 to 6.7 percent of gross domestic product (GDP) for a 3°C increase in global average surface temperature (Nordhaus and Moffat 2017; Howard and Sterner 2019; Tol 2018); see Figure 1. However, these two damage functions were both estimated using the meta-analysis technique, a method long considered the objective and scientifically rigorous way for combining results from multiple studies to develop consensus (by improving the overall estimate of treatment effect size) and resolve uncertainties and discrepancies within the literature (by explaining these differences) (Stegenga 2011; Holman 2019). Since the damage function is so very important for the calculation of an optimal climate policy and since such large discrepancies remain despite similar methodologies, we believe that it is vital to further refine our methodology.

The wide discrepancy between Nordhaus and Moffat (2017) and Howard and Sterner (2017) may not be surprising, given the significant uncertainties underlying climate damage functions (Nordhaus 2019). In the United States, the social cost of carbon (SCC) is a critical climate policy tool. However, the difference in meta-regression damage estimates implies a 3.5- to 5.5-fold difference in the 2020 SCC estimate within DICE-2016R2 and a 0.7°C to 0.9°C difference in the optimal global average surface temperature in 2100. Moreover, given the recent burst of climate damage publications (Greenstone 2017), the use of meta-analysis to synthesize this growing literature at the sector (e.g., Moore et al. 2017) and global levels (e.g., Tol 2009) is becoming increasingly popular. In this paper, we analyze differences in meta-analysis results to determine what causes this discrepancy and whether it stems from inconsistency in the application of methodologies or from the subjectivity inherent in the summary of climate damage literature. Using this understanding, we propose an improved estimator adapted from the medical literature.

Finding drastically different results from earlier summaries by Tol (2009, 2014) and Nordhaus and Sztorc (2013), Howard and Sterner (2017) explained many of the discrepancies. Specifically, Howard and Sterner (2017) introduced up-to-date regression techniques and methods (some from the medical sciences) into the global damage context to address ongoing issues of systematic review, duplicate publication bias, heteroskedasticity, lack of methodological controls, measurement error, dependent errors, and outliers. The authors found that the use of differing sets of damage estimates greatly influenced the meta-analysis results, and the importance of duplication bias and omitted methodological controls depended on whether high-temperature damage estimates ( $\geq 4^\circ\text{C}$ ) were excluded. Nordhaus addressed some of

these issues in his latest analysis by introducing a systematic review synthesis (SRS), outlier robust methods, and estimate-based weights (Nordhaus and Moffat 2017). However, as noted by Nordhaus (2019), the discrepancy remains. Some of this difference still results from inconsistent application of meta-analysis techniques, most prominently methodological controls and heteroskedasticity, as raised in Howard and Sterner (2017). Additional reasons may be that Nordhaus and Moffat (2017) introduce author-determined estimate-specific weights along with several other methodological differences; see Table 1.

Our analysis finds several methodological causes for the discrepancy in meta-analysis results, though differing subjective opinion is also a driving factor. On the one hand, omitted methodological controls (particularly controls for the inclusion of productivity, nonmarket, and catastrophic impacts in addition to the more traditionally captured direct market effects) and the failure to address heteroskedasticity are major methodological reasons for the differences in results between Howard and Sterner (2017) and Nordhaus and Moffat (2017). Although duplication is still an issue, we no longer find evidence that this is the driving methodological cause of the difference. On the other hand, in addition to the search and selection rules, the inclusion of explicit and implicit weighting of estimates is another major source of discrepancy. To avoid implicitly weighting estimates at zero, an improved systematic review process that balances the identification of relevant studies and the screening needs for assessors is critical. Although improving search rules is beyond the scope of this paper, we develop transparent rules for explicit weighting of global climate damage estimates and apply a recently developed quality effects estimator designed to address the subjectivity of weights finding climate damages of 7 to 16 percent for a 3°C increase, depending on the scope of climate impacts considered. Because eliminating subjectivity from this process may be impossible, future work should consider using expert elicitation to develop weights for estimating a central (i.e., “consensus”) damage function.

## **2. Differences between Nordhaus and Moffat (2017) and Howard and Sterner (2017)**

To reflect advances in climate science and economics, Nordhaus and Sztorc (2013) updated the DICE global damage function by switching its calibration from the enumerative approach to a meta-analysis. Using global damage estimates assembled by Tol (2009), in the first meta-analysis of global climate damage, Nordhaus and Sztorc (2013) regressed temperature squared on damages and then adjusted the coefficient upward by 25 percent to account for unmeasured nonmarket and catastrophic impacts. After two sets of corrections to the Tol (2009) data set underlying the DICE-2013R damage function (Tol 2014; JEP 2015), Nordhaus

conducted his own SRS and meta-regression of global damage estimates (Nordhaus and Moffat 2017; Nordhaus 2019). After correcting significant errors in the Tol (2014) data, adding several additional damage estimates (via SRS), and improving the econometric techniques applied, including the application of median regression and quality weights, the DICE-2016R damage function reestimated in NM declined by 17 percent relative to the 2008 model.<sup>1</sup>

Around the same time, Howard and Stern conducted their own review of the global damage literature. After similarly correcting previous data errors and adding a variety of damage estimates, HS dropped a large subset of estimates for not meeting the selection criteria and to address estimate duplication. Using this data set, HS estimated an alternative DICE damage function using updated economic techniques to address omitted variable bias, heteroskedasticity, dependent estimates, measurement error, and outliers. In contrast to NM, the HS damage function is a 150 to 300 percent increase relative to DICE-2008, depending on the inclusion of productivity and/or catastrophic impacts.

To explain this discrepancy, this section reviews the differences in the data and estimation methodologies applied by NM and HS. Table 1 summarizes these differences.

## 2.1 Data: Search, Selection, and Weighting

Unlike prior studies, HS and NM attempted to be systematic in their data collection and selection. Before running the meta-regression, Nordhaus and Moffat conducted a system review synthesis. However, they concluded that a formal process such as SRS would be unsuccessful in the context of “climate change impacts” because of the diffuse nature of the field. Specifically, they found that formal search procedures generated either insufficient data (failing to capture known studies in Tol 2014) or too many studies (of which few were useful). Instead, the authors applied a combination of formal search via EconLit and informal methods to identify the relevant set of observations. For similar reasons, Howard and Stern also used a variety of formal and informal methods, which they rigorously documented, including applying a combination of search algorithms in Google Scholar and EconLit, looking at climate change publications of already-included authors, and including studies cited in up-to-date meta-analyses of the SCC and climate impacts.

---

<sup>1</sup> Using similar data and differing methods, Tol (2018) recently found a nearly identical damage function as DICE-2016R when applying Nordhaus’ quadratic damage function. However, we focus on comparing the results of HS with NM, instead of Tol (2018). As little has changed in Tol’s methodology between earlier versions of his model and the 2018 version, HS already explains much of the difference between HS and Tol (2018); see Section SM.1 in the Supplementary Material for more discussion.

Given the informal portion of both searches, we cannot determine exactly which papers were identified for review by these differing search methods. By design, both NM and HS include all Tol (2014) papers in their respective data pools. However, remaining differences in their formal and informal search processes likely lead to different study pools from which to select relevant data. Using the search criteria provided by Nordhaus for EconLit, we can identify several papers in HS that must have been in the NM pool (and were later dropped) via the formal process: Ackerman et al.'s (2012) application of the Hanemann (2008) and Weitzman (2012) impact estimates and an early version of the MERGE model (Manne et al. 1995). Additionally, certain important papers in the HS pool—including Hanemann (2008), Dell et al. (2012), Weitzman (2012), Nordhaus (2014), and Burke et al. (2015)—were likely included in the NM pool (and later dropped) via the informal process because of the prominence of the papers and authors in the literature.<sup>2</sup> Several other papers in HS appear not to be in the NM pool: Meyer and Cooper (1995), Schauer (1995), Horowitz (2009), Bluedorn et al. (2010), Ng and Zhao (2011), and Howard and Sylvan (2015); several of these are from the gray literature.<sup>3</sup> Similarly, using the HS search criteria, we can identify all estimates in NM that were in the HS pool (and were later dropped) via the formal process: Cline (1992); Kemfert (2002), Nordhaus (2010), Dellink et al. (2014), and Hambel et al. (2015). Given the wide net cast by the HS search algorithm using Google Scholar, the lesser-known studies were analyzed for relevance relatively quickly in the selection process.

Like the search process, the selection criteria slightly differed between the two studies, leading to only partially overlapping data sets; see Table SM.1 in the online Supplementary Material. NM applies a broad selection criterion that studies must supply an estimate of the “global impacts of climate change” and then applies weights to these estimates. Because this wide criterion includes willingness-to-pay and willingness-to-accept estimates and output (usually GDP) and consumption-based estimates, it is unclear how any studies selected by HS, if in the NM pool, would have been rejected. Likely several were screened out by NM, which implicitly assigned a weight of zero to these estimates, including all those drawn from time-panel (“weather”) studies (Dell et al. 2012; Burke et al. 2015) and scientific studies (Weitzman 2012; Nordhaus 2014), based on their perceived low quality.

Despite casting a wide net in the search process, HS applies narrower selection criteria: estimates must capture “global willingness to pay to avoid climate change measured as a percentage of global GDP” and must not be duplicates (i.e., superseded, derivative, updated, or multiples), must not arbitrarily cap damages based on author

---

<sup>2</sup> Additionally, NM appears to reference indirectly the work of Dell et al. (2012) and Burke et al. (2015): “our opinion is that studies relating current incomes to ... weather provide little information about long-run impacts of climate change, so these receive a low weight.”

<sup>3</sup> Some of the papers in this gray literature have been published since 2017, including Howard and Sylvan (2020).

discretion, and must not fail to address omitted variable bias in statistical contexts.<sup>4</sup> Based on these selection criteria, none of the studies added through the NM research synthesis (Cline 1992; Nordhaus 2010, Dellink 2012, Kemfert 2001, and Hambel et al. 2015) would be included in HS because of their duplication, geographic extent (e.g., Cline 1992), and omitted data (e.g., Hambel et al. 2015, 2021); it appears that Cline's (1992) US damage estimates should also be excluded from NM as well, based on the global selection criterion.<sup>5</sup> From Tol (2009), HS also drops three compensating surplus estimates by Maddison and Rehdanz that capture only nonmarket effects (Maddison 2003; Rehdanz and Maddison 2005; Maddison and Rehdanz 2011) because they are not strictly willingness-to-pay estimates. Partially because of its differing treatment of duplicate estimates, NM has more damage estimates (36 estimates from 27 studies) than HS (26 estimates from 20 studies).

The two papers also take different approaches to weighting; see Table 1. To address lower informational content from duplication and poor quality, NM uses weights to focus the meta-analysis on the latest version of independent studies that apply "appropriate" estimation methods. To studies that do not meet these criteria (i.e., they are superseded or derivative or they apply poor methods), the authors assign a 10 percent weight.<sup>6</sup> Though no specific time element is discussed, earlier and updated versions of studies are assigned a weight of 25 percent, versus a 0 percent weight for outdated studies and a 10 percent weight for "superseded" studies. Weights are assigned to multiple estimates from the same study such that they sum to one, with greater weight placed on less speculative and central estimates. In contrast, HS down-weights estimates by the inverse of their temperature change (as a proxy for estimate precision) to address heteroskedasticity, and drops (i.e., assigns a weight of zero to) estimates greater than 4°C in their preferred specification to address their speculative nature (implicitly, HS assumes that a strong nonlinearity arises above 4°C in the level of damage uncertainty, causing the fixed effects weight to collapse to zero).<sup>7</sup> Inevitably, the application of explicit weights interjects an element of subjectivity into both studies, as noted by NM, though such subjectivity or judgment is also present in the search and selection steps (e.g., the decision to drop a study could be seen as an implicit weight of zero).

After implementing the above three steps, the two studies have different data sets; see Table 2. Most notably, the ratio of damages to temperature increase is higher for HS than for NM, except when duplicates are dropped and weights applied to both data

---

<sup>4</sup> Duplication plagues earlier meta-analysis of climate damages; see Figure 2 in HS for an example.

<sup>5</sup> See Section SM.2 in the Supplementary Material for further discussion.

<sup>6</sup> NM do not explain their reason for selecting the magnitude of their weights.

<sup>7</sup> NM also drop estimates based on temperature criteria in their sensitivity analysis to address outliers, though not in their preferred specification.

sets. Because the methodological variables are only slightly different,<sup>8</sup> this difference is likely driven by the types of estimation methodologies and studies included and favored in the underlying data set. In particular, the NM data set is heavily weighted toward enumerative estimates, which, together with computable general equilibrium (CGE) models receive 76 percent of the weight in NM's preferred specification, compared with 30 percent in HS. Instead, the other three general estimation methodologies receive more weight in HS, including statistical studies, which receive the bulk of the weight, and natural science-based ("scientific") estimates, which receive zero weight in NM.<sup>9</sup>

The NM and HS weights also place different emphases on established authors. NM's preferred specification places 90 percent of its weight on estimates from Yale or the University College of London groups (see Tol 2009), compared with 40 percent in HS. Specifically, Nordhaus's own estimates receive almost 40 percent of the weight in NM, compared with only 15 percent in HS.<sup>10</sup> Nordhaus, Tol, and Bosello jointly receive a 75 percent weight in NM, compared with 27 percent in HS. Finally, HS places slightly more than a quarter of the weight on the gray literature compared with NM, which includes no gray literature at all.

Weighting of estimates is an important issue in the literature: the estimation strategies are nonrandom because certain methodologies are more common, and the magnitude of effects depends on the underlying estimation methodology (Howard and Sterner 2017). Therefore, the implicit weighting of different methodologies (i.e., their relative weights in a summary) may greatly determine the results. Like estimation strategies, publication in high-ranking journals is not random: established authors' studies are more likely to be published, and outlier estimates less so (i.e., publication bias). Thus, more established authors using older methods (enumerative and cross-sectional studies) have more publications.<sup>11</sup> It is easy to see that NM weights more toward older studies, with an average publication year before 2006 for NM versus after 2008 for HS. However, because enumerative and CGE studies are based on earlier studies, predominately from the 1990s and early 2000s (Greenstone 2017), this understates this difference. Thus, NM implicitly and explicitly puts 82 to 98 percent of the weight

---

<sup>8</sup> The exception is that NM includes a higher share of studies with nonmarket effects due to the inclusion of the compensating surplus estimates (Maddison 2003; Rehdanz and Maddison 2005; Maddison and Rehdanz 2011).

<sup>9</sup> See Section SM.3 for a discuss of these estimation methodologies and their relative advantages and disadvantages.

<sup>10</sup> This increases to 23 percent in HS with the inclusion of damage estimates above 4°C.

<sup>11</sup> The publication of a study in a high-ranking journal is likely correlated with high quality, though not perfectly. For instance, established authors might be overrepresented in high-quality journals even after taking the high quality of their papers into account. These same authors might also rely on older, more established estimation methodologies.

on damage estimates from 2006 or earlier, versus 36 to 49 percent in HS.<sup>12</sup> Hence, meta-analysis techniques need to address other nonrandom elements that weight climate damage meta-analyses in one direction or another.

## 2.2 Methodological Differences

To develop their estimators, HS and NM start with different assumptions about methodological effects on the damage estimates. In Nordhaus's initial meta-regression, Nordhaus and Sztorc (2013) ignore the hierarchical nature of meta-data and implicitly assume a true relationship between temperature and damages that varies *only* with temperature squared and random error. The resulting econometrics model is relatively simple:

$$(1) D = \alpha T^2 + \varepsilon \text{ where } \varepsilon \sim N(0, \sigma^2).$$

where  $D$  is a study estimate of global climate damages (measured as a percentage of global GDP),  $T$  is global average surface temperature and is assumed to be the sole factual cause of observed heterogeneity,<sup>13</sup> and  $\varepsilon$  is the error term. NM essentially maintains the above model, introducing quality weights partially based on poor estimation quality in recognition that estimation methodology has a discernible effect on the relationship between temperature and damages. Even so, NM ignores the hierarchical structure of meta-analysis stemming from the uncertainty over the underlying damage estimates, as shown below.

In contrast, HS recognizes this hierarchical structure and also assumes a distribution of true temperature-damage relationships because methodological differences between studies result in estimates that capture varying portions of damages. HS develops a model that captures the methodological causes of variation, which partially stem from the variety of methods for estimating climate damages in the literature; it also captures clustering due to dependency between studies with similar authors, estimation methods, and models. Specifically, their econometric specification is

$$(2) D_{i,j,k} = \alpha t_{i,j} + \sum_l \alpha_k t_{i,j} X_{i,j,k} + \varepsilon_{i,j,k} \text{ where } \varepsilon_{i,j,k} = \mu_i + e_{i,j} \text{ and } \mu_i \sim N(0, \tau^2) \text{ and } e_{i,j} \sim N(0, \sigma_{i,j}^2)$$

---

<sup>12</sup> Because a handful of studies that underlie DICE, FUND, and PAGE were published after 2006, we calculate these ranges by assigning a weight of two-thirds to one to enumerative and CGE estimates for assigning these estimates to the pre-2006 category.

<sup>13</sup> Beyond historical precedent, both NM and HS assume a quadratic damage function to ensure that the damage function increases at an increasing rate and is relatively stable, while also preserving degrees of freedom.

where  $t_2$  is global average surface temperature increase relative to the base period of the study squared,  $j$  is the cluster level,<sup>14</sup> and  $X_{i,j,k}$  is an indicator variable equal to one if estimate  $i$  is characterized by methodological factor  $k$ . Given the hierarchical structure, the error term ( $\varepsilon$ ) can be subdivided into unobservable methodological differences ( $\mu$ ) and measurement errors ( $e$ ) where  $\tau^2$  equals damage heterogeneity unexplained by temperature squared and its interaction with observed methodological differences (residual heterogeneity or residual between study variation) and  $\sigma_i^2$  is the variance underlying damage estimate  $i$  (Harbord and Higgins 2008). Because HS finds that methodological controls are significant, and because estimates that capture differing climate impacts can be directly observed in their construction, the HS model does not collapse to the NM model.

Specifying different econometric models leads to the selection of different estimators. Based on its simple model, NM uses weighted least squares and weighted median regression (to address outliers), with weights solely designed to address duplication and methodological quality. However, in addition to the issue of omitted variable bias, Expression (2) highlights that the resulting coefficient estimates in the NM model are likely to be inconsistent and inefficient and the corresponding standard errors are biased by the common issues of heteroskedasticity and dependent errors.<sup>15</sup> To address heteroskedasticity, Howard and Sterner (2017) apply the standard fixed effects estimators—a version of weighted least squares where the precision-based weights are  $\frac{1}{\sigma_{i,k}^2}$  (US EPA 2006; Nelson and Kennedy 2009; Stanley et al. 2013). Implicitly, HS assumes that  $\tau = 0$ , implying that the fixed effects model is the best linear unbiased estimator, since otherwise the random effects estimator—a version of weighted least squares where the precision-based weights are  $\frac{1}{\sigma_{i,k}^2 + \tau^2}$ —is traditionally preferred.<sup>16</sup> This assumption is necessary because  $\sigma_i^2$  is observable for only a handful of the underlying studies. HS instead runs the fixed effects model, adopting the inverse of the temperature change applied in the underlying study as a proxy measurement of the precision weights (i.e.,  $\frac{1}{t} \approx \frac{1}{\sigma_{i,k}^2}$ ).<sup>17</sup>

HS finds that the methodological controls are significant implying that NM suffers from omitted variable bias. In fact, NM's failure to include methodological controls partially explains NM's flatter relationship between temperature change and climate damages.<sup>18</sup>

<sup>14</sup> In HS, Howard and Sterner cluster at the model level, which is defined by the damage estimation method and primary author.

<sup>15</sup> Due to the sample size in NM, lost consistency and efficiency is particularly problematic for the resulting estimates.

<sup>16</sup> The implicit assumption that  $\tau=0$  is also reasonable given that HS controls for the main methodological differences between studies, such that residual heterogeneity should be small.

<sup>17</sup> This is similar to the standard strategy of using the number of observations in each study to proxy for inverse variance, as variance declines with the number of observations (Nelson and Kennedy 2009).

<sup>18</sup> These results are stable to a variety of robustness checks, including dropping potential outliers (above 4°C) and outlier-robust estimators.

Specifying different econometric models, particularly the inclusion of methodological controls, also leads to multiple measurements of climate damages. In Expression (1), NM does not address the fact that climate studies differ in the climate impacts they capture, such that NM treats each climate damage estimate as a random draw from the same distribution. However, this is not accurate: various estimation strategies capture differing types of climate impacts. Although almost all climate studies capture direct effects of climate change on market sectors, only certain empirical methodologies for estimating climate damage can capture productivity effects (statistical and CGE estimates), nonmarket impacts (enumerative, survey, statistical, and scientific), and catastrophic risk premiums (enumerative and scientific)<sup>19</sup>; see Section SM.3 in the Supplementary Material for further discussion. In contrast, HS includes  $X_{i,j,k}$  as a set of indicator variables equal to one if estimate  $i$  includes only market effects, includes catastrophic impacts, accounts for productivity (i.e., market effects via productivity shocks), and corresponds to a cross-sectional estimate without country fixed effects. Specifically, the HS model is

$$(3) D_{i,j} = \alpha t_{i,j} + \alpha_1 mkt\_t_{i,j} + \alpha_2 cat\_t_{i,j} + \alpha_3 prod\_t_{i,j} + \alpha_4 cross_{i,j} + \varepsilon_{i,j,k}$$

where  $\alpha$  is the direct market and nonmarket effects of climate change;  $\alpha + \alpha_1$  is the nonmarket effects;  $\alpha_2$  captures the catastrophic impacts; and  $\alpha_3$  captures the productivity consequences of climate change. Accordingly, HS captures four measurements of climate damages: noncatastrophic ( $\alpha$ ), total ( $\alpha + \alpha_2$ ); noncatastrophic, including productivity shocks ( $\alpha + \alpha_3$ ), and total, including productivity shocks ( $\alpha + \alpha_2 + \alpha_3$ ).

---

<sup>19</sup> Only the statistical studies using time-panel data capture productivity effects, and only cross-sectional studies using happiness as the endogenous variable and the household production function approaches capture nonmarket effects. In fact, these happiness and household production function studies are the only studies that do not capture market effects of climate change, and they are excluded from HS because they are not willingness-to-pay estimates.

### 3. Analyzing Regression Differences

In this section, we formally analyze the exact cause of the discrepancy between the HS and NM meta-regression results. We start by replicating the NM results using the HS data. Specifically, we start with the HS regression using the HS data and demonstrate the step-by-step changes necessary to replicate the NM results: (1) include data above 4°C; (2) drop methodological controls and scientific estimates; (3) drop panel estimates; (4) drop estimates from the gray literature; (5) drop adjustments to temperature to account for the base period of the study and clustering; (6) add missing NM data and drop HS's inverse temperature weights; (7) add NM weights and switch Nordhaus (1994b) and Tol (2013) to reflect NM's reading of the estimates; and (8) adjust weights to replicate equal weighting of estimates. This is followed by an analysis of these same assumptions one by one instead of collectively for HS and NM, both with and without damage estimates above 4°C. Finally, we explore the effect of outliers.

#### 3.1 Replication of Nordhaus and Moffat (2017)

Table 3 captures the step-by-step changes in HS's assumptions to NM's assumptions to replicate the NM results. Specifically, we implement each step consecutively (read cumulatively, from left to right), starting with HS's preferred specification in Specification 0 and ending with NM's weighted and unweighted ordinary least squares (OLS) estimates in Specifications 7 and 8, respectively; our approximations of NM's damage estimates corresponding to Steps 7 and 8 are within -2 and 11 percent of the originals, indicating that our methodology represents a good approximation of NM.

Our successful replication highlights that NM duplicates several damage estimates. Except for importing the Kempfert (2002), Cline (1992), and compensating surplus estimates from Maddison and Rehdanz (2011), all other missing NM estimates added in Step 6 are replicated through reweighting HS papers to account for duplication in the NM study; see Table SM.1. Specifically, to account for duplicate studies that we exclude, we increased the weights on Nordhaus (2008a), Tol (2013), and Bosello and Parrado (2014) to above one, giving them values of 2.7, 1.8, and 2.5, respectively.<sup>20</sup> However, we find that this duplication does not significantly affect the NM results.<sup>21</sup>

---

<sup>20</sup> Additionally, we assign a weight of zero to any remaining published estimates cited in HS that NM excludes. As such, our version of the NM weights in Specification (7) sums exactly to 12.25, which is the same aggregate weight assigned in NM. Furthermore, when we assume equal weighting in Specification (8), we adjust the weighting to sum to 38, which is the number of observations in NM. Thus, the same relative and absolute weights hold as in NM.

<sup>21</sup> Using our NM data set and weights, we run multiple tests for duplication bias and find little to no evidence. The lack of duplication bias in NM is likely the result of down-weighting estimates

The primary driver of difference in this analysis is different weighting procedures. The three largest are, in order, Step 1, which includes high-temperature estimates by removing the zero weighting for estimates above 4°C in HS (this step shifts the noncatastrophic damage component to the catastrophic damage component); Step 7 of applying the NM weights; and Step 3, which drops all time-panel estimates (which is equivalent to NM's 0 percent implicit weighting of “weather” studies). To put these changes in perspective, switching from weighted OLS regression (Specification 7) to a median regression in NM decreases the estimate results by only 15 percent, whereas all three changes far exceed this effect in percentage terms. Surprisingly, controlling for methodological variables (i.e., Step 2) has little effect, though it may be partially a result of the order of changes made.

### 3.2 One-by-One Regressions Analyzing HS

To account for order of operation, we isolate the effect of each of these earlier steps (i.e., operations) in Subsection 3.1 by conducting each step separately on the HS preferred regression specification (i.e., estimates below 4°C) and all damage estimates. Using the HS data, we also conduct a similar analysis removing each control variable from the preferred HS specifications. Finally, we conduct a similar order-of-operations analysis on the NM analysis, though this is more limited. See Tables SM3–5 in the Supplementary Material.

The regressions using the HS data further highlight the importance of weights, since including damage estimates above 4°C and including NM weights still significantly affect the results (as discussed above), along with dropping the HS weights. However, excluding time-panel data now affects only the coefficient corresponding to productivity, since control variables are still included in the regression. In contrast, removing control variables, in particular the control variables for market-only effects and catastrophic impacts, now greatly affects the damage estimate as captured by the coefficient corresponding to temperature squared (our primary variable of interest that captures direct market and nonmarket damages). Because the catastrophic variable absorbs much of the high-impact estimates, the inclusion of science-based estimates is not driving the difference between HS and NM.<sup>22</sup>

Using the NM data, we conduct a similar analysis focusing on the two main assumptions: weights and control variables. However, since the NM, HS, and the above analyses show that global damage meta-regressions are relatively robust to individual modeling changes, we apply individual and multiple modeling changes to the NM regression, similar to Section 3.1. After removing duplicate estimates, we explore the effect of including methodological variables when dropping the sole estimate above

---

that are outdated or drawn from superseded earlier studies that have been updated by the authors.

<sup>22</sup> Dropping gray literature has a noticeable effect on damage coefficients, though HS found no evidence of publication bias.

8°C (i.e., the Cline 1992 damage estimate of 6 percent of GDP for a 10°C increase with a 10 percent quality weight), following NM, and then excluding all estimates above 4°C, following HS. Excluding the “highly speculative” Cline (1992) estimate alone increases climate change damage estimates by at least 61 percent. Excluding the three studies above 4°C further increases climate damages, even exceeding the HS estimate when control variables are included. Thus, the inclusion of control variables and the treatment of highly speculative damage estimates corresponding to high-temperature increases can be consequential.

### 3.3 Outliers

Both NM and HS address outlier estimates, dropping estimates based on temperature thresholds. Whereas HS drops estimates above 4°C in their preferred specification as too speculative (i.e., HS speculates that the variance of the damage estimates increases rapidly above 4°C, leading to a near-zero weight in the fixed effects estimator), NM conducts a series of runs dropping estimates below and above various temperature thresholds.<sup>23</sup> In addition, HS reruns the model, dropping each estimate and study to determine its individual effect on the results. Both papers also run outlier robust estimators. Whereas NM’s preferred model is a median regression model, HS is critical of this approach because median regressions address outliers in the dependent variable space (damages), not the independent variable space (temperature).<sup>24</sup> Instead, HS runs multiple outlier robust estimators, including the efficient MM-estimator. Using the above methods, HS identifies several possible outlier estimates, though they are dependent on the model specification and whether asymptotic properties hold with the small HS data set.<sup>25</sup>

We rerun this analysis for NM and HS using MM-regression to identify outliers, finding that high-temperature damage estimates ( $\geq 4^\circ\text{C}$ ) are frequently outliers; see Table S6.<sup>26</sup> Using the weighted NM data, we find that both Cline (1992)’s estimates at 2.5°C and 10°C and Maddison and Rehdanz (2011)’s estimate at 3.2°C are potential outliers. Using the inverse-temperature weighted HS data, we find that Weitzman (2012)’s estimate at 6°C, Burke et al. (2015)’s estimate at 4.3°C, and Howard and Sylvan (2015)’s estimate at 3°C are potential outliers. Several high-quality estimates according to the NM data are outliers only when control variables are included, including Nordhaus (1994b)’s estimates at 3°C and 6°C and Fankhauser (1995)’s

---

<sup>23</sup> It is unclear why NM test sensitivity to low and high temperatures, since high-temperature estimates are clearly more speculative.

<sup>24</sup> NM address this shortcoming through combining their median regression model with temperature thresholds.

<sup>25</sup> Several studies are potential outliers: Meyer and Cooper (1995), Schauer (1995), Nordhaus (2008a), Horowitz (2009), Weitzman (2012), Nordhaus (2014), Burke et al. (2015), and Howard and Sylvan (2015).

<sup>26</sup> We include the weights as probability weights instead of analytical weights because of the limitations of *robreg*.

estimate at 2.5°C,<sup>27</sup> while this is also true for Nordhaus (2014)'s 3°C estimate for a strictly binding 2°C target and Meyer and Cooper (1995)'s 3°C estimate when controls are included. From these results, we can see that speculative high temperature damage estimates ( $\geq 4^\circ\text{C}$ ) and less frequently used estimation methodologies (scientific-based, survey, and time-panel data) tend to be outliers. Using the MM-estimator, the NM and HS results are relatively robust to outliers. While our NM results are equivalent to NM's preferred median regression, we find that accounting for outliers in the HS model mostly just shifts impacts from non-catastrophic (decreasing by -27% to -32%) to catastrophic (increasing by 4% to 38%), such that total impacts decline only slightly (by -12% to -14%) when using the HS data. Even when addressing outliers, the HS estimates are generally higher than the NM estimates, particularly when HS drops high-temperature damage estimates and/or accounts for catastrophic impacts, though the disparity is not as wide. Even so, care should be taken when interpreting these results, since outlier status appears related to the potential informational content of data and dependent on identifying and selecting all relevant data.<sup>28</sup>

Like most of the sensitivity analysis in HS, including speculative high-temperature estimates (above 4°C) and addressing outliers shifts noncatastrophic impacts to catastrophic. However, this result is partially driven by the general classification of damage estimates as catastrophic in HS, combining risk premiums, catastrophically high temperatures, and catastrophically high draws from the underlying damage distribution. Narrowing the definition of catastrophic to reflect solely the economic concept of risk or insurance premium corresponding to catastrophic risk, as applied by Nordhaus (2019) in DICE, changes the indicator variable for catastrophic impacts. Whereas DICE-1999 and DICE-2008 explicitly include a risk premium and Nordhaus's (2014) scientific damage estimates are clearly based on scientific concerns of triggering tipping points from crossing the threshold of 2°C (i.e., they reflect a risk or insurance premium), Weitzman's (2012) estimates are catastrophic in the sense that losses are large from experiencing high temperatures. Since the Weitzman estimates are based on Sherwood and Humber's (2010) work on the limits of the human respiratory system globally, Weitzman's 6°C estimate should be weighted less than his 12°C estimate, which is more soundly based in science. Similarly, Meyer and Cooper's (1995) willingness-to-pay estimate was classified by HS as catastrophic based on Fankhauser's and Tol's criticism of their malnutrition estimate as excessively high (i.e., catastrophic draw from the damage distribution), though this estimate should instead be down-weighted to represent its poor quality, in addition to being reclassified as noncatastrophic. After adjusting the catastrophic classification of Weitzman (2012) and Meyer and Cooper (1995) and down-weighting the poorer-quality estimates (see

---

<sup>27</sup> In the NM dataset, Kempfert's (2002) estimate for a 0.3°C increase is also an outlier when controls are included.

<sup>28</sup> Specifically, high-temperature damage estimates and less frequently applied methodologies provide more information and are frequently identified as outliers. Furthermore, Nordhaus (1994b) is not an outlier in the HS data set but is in the NM data set, indicating the importance of data context.

Table SM.7), HS's damage estimates become insensitive to the inclusion of damage estimates above 4°C, though the pattern of shifting from noncatastrophic to catastrophic impacts continues for the robust estimator (along with a decline in total damages overall). Whether to define catastrophic in more general or more narrow terms is relatively subjective: both definitions are acceptable if the assumption is transparent, though this newer definition is more relevant in the context of deterministic runs in DICE.

### 3.4 Takeaways

From the above analysis, we can see that the HS and NM results differ for both objective and subjective reasons. In the former category, we found that the inclusion of methodological control variables, in particular the inclusion of the market-only and catastrophic control variables, can affect the results. Additionally, the use of a weight to address heteroskedasticity (i.e., an inverse temperature weight) further deemphasizes speculative and highly uncertain high-temperature estimates ( $\geq 4^\circ\text{C}$ ), which tend to be outliers. This down-weighting is critical because the meta-regression results are sensitive to these outliers in the temperature space, such as the 10°C estimates from Cline (1992).

In the latter category, we found that the selection of damage estimates and choice of quality weights greatly affect the results, partially by assigning implicit weights to the various estimation strategies employed in the climate damage literature. For example, NM's weights fall almost completely on the enumerative and CGE methods (and estimates by established authors) and less so on other identification methodologies weighted more by HS (scientific, statistical, and survey). Consequently, HS results are nearly identical to NM when we simply weight HS's enumerative and CGE methodology specific estimates (i.e., 0.212 for enumerative only and 0.151 for CGE only) by the corresponding NM weights: 0.196 compared with 0.173. This is problematic because the number of estimates from each method is nonrandom, and the temperature-damage relationship differs by the estimation methodology (see Tables C9 and C10 in HS). This also raises concerns about outlier robust estimators, since these estimators may be simply dropping estimates using more recent techniques underrepresented in the data set. However, the inclusion of scientific and time-panel estimation methodologies by HS, which are excluded from NM, is not driving the difference between the studies, partially because of the inclusion of control variables by HS (Howard and Sterner 2017). Even so, the definition of catastrophic drives the disparity between the HS estimates that exclude and include damage estimates above 4°C.

## 4. Improving Meta-Analysis: Addressing Methodological and Data Discrepancies

The results in Section 3 show the objective (i.e., methodological) and subjective (i.e., data search, selection, and imputation and quality weight) reasons for the difference between the HS and NM results. In this section, we focus on the data search, selection, and entry discrepancies and methodological assumptions from HS and NM; we leave quality weighting to the following section, to introduce new concepts.

In data assembly, some level of subjectivity is inevitable. Therefore, analysts should try to minimize it by formalizing processes and, where those processes cannot effectively reduce subjective bias, by employing transparency, consistency, and sensitivity analysis. This is true to a lesser extent with methodological decisions, though transparency and sensitivity analysis are still best practices to ensure the robustness of findings. In this section, in making our preferred data and modeling decisions, we err on the side of Nordhaus on subjective assumptions to avoid driving our modeling results toward HS.

One way to reduce subjectivity in meta-regressions is to apply formal techniques to data search methods. For example, NM attempts to implement an SRS, but as the authors note, the diffuse nature of the climate change field makes a formal search process difficult: studies estimate different effects, apply different estimation methods and identification techniques, and measure social consequences differently. A potential solution employed by both NM and HS is to combine an informal process with a formal one subject to rigorous documentation. However, subjective decisionmaking nevertheless enters the search process, leading to substantially different data sets. Future work should engage with data scientists to improve the formal search process and ensure a balance between screening out irrelevant studies and capturing relevant work. Since this is beyond our scope here, we combine the NM and HS data sets to construct our data pool.

Given also the potential subjective bias in the selection and data entry processes, researchers must be transparent and document their decisions. Researchers must determine their selection and data entry criteria *a priori* to reduce bias (Howard and Sterner 2017), implement their selection and data entry procedures rigorously, and document difficult decisions. Because of the subjective element of these criteria themselves, researchers should conduct sensitivity analysis when multiple assumptions can be justified.

Consistent with those best practices for selection, we use the NM criterion of all global damage estimates. Deviating from NM, however, we follow HS in dropping duplicate

estimates—the methodologically correct approach, in our opinion. It is also difficult to determine whether NM and HS captured all earlier and superseded versions of the models in their data sets (Howard and Sylvan 2017). Rigorously implementing the NM selection criteria, we include all relevant HS data and include all compensating surplus nonmarket damage estimates from Maddison and Rehdanz (2011) excluded in HS, and Kemfert (2002) from NM. Critically, we exclude all duplicate data and US-only estimates, including Cline (1992). We conduct sensitivity analysis for the treatment of duplicate estimates, the focus on willingness-to-pay studies, and the cutoff on what makes a global versus domestic estimate.

For data entry, we mostly follow the HS definitions, since NM did not categorize the types of climate impacts captured by each damage estimate. However, we make two exceptions differing from HS, regarding the definition of catastrophic impacts and estimate quality. Moving forward, we adopt the narrower definition of catastrophic impacts as risk or insurance premium (discussed in Section 3.3). In our opinion, this is the appropriate definition for applying the resulting damage function in a deterministic integrated assessment model (IAM), like DICE.<sup>29</sup> Because HS does not assign estimate quality, except to implicitly assign weights of 0 percent to estimates exceeding 4°C, we adopt NM’s definition of poor quality, including for happiness and weather-based damage estimates (in this section) despite our differing opinion, and their decision to weight poor-quality estimates at 10 percent. For data not considered by NM, we categorize them ourselves. Based on our earlier discussion, we categorize Meyer and Cooper (1995) as poor quality. Additionally, although we assign 100 percent weight to the scientific studies (Nordhaus 2014; Weitzman 2012), each of which produces two damage estimates, we assign a weight of 90 percent to the more scientifically sound estimate and 10 percent to the other.<sup>30</sup> We conduct sensitivity analysis for the definitions of “catastrophic” and “poor quality.”

Methodologically, we rely on the HS model, which we believe is objectively preferred. Thus, we include methodological variables for market-only, catastrophic, and productivity impacts, as discussed earlier. We deviate from HS’s preferred specification in that we include a compensating surplus control variable to account for differences in compensating surplus and willingness to pay. We apply the fixed effects estimator, weighting estimates by their inverse temperature, following HS. However, because we also include quality weights, we use the product of the two weights in practice. To address dependent errors, we cluster errors at the model level, following

---

<sup>29</sup> Additionally, the HS damage estimates are robust to the exclusion of damage estimates above 4°C under this narrow definition of catastrophic (see Section 3.3).

<sup>30</sup> Weitzman’s (2012) 6°C estimate is of lower quality than the 12°C estimate because Huber and Sherwood (2010) do not model human migration away from inhospitable locations. However, this is less of an issue for the 12°C estimate, since at this point, most of the planet is inhospitable. Nordhaus (2014) estimates climate damages if 2°C is strictly optimal or optimal with overshoot. Since most scientists appear to refer to a strictly binding target as ideal, we assume that overshoot represents a lower-quality estimate.

HS. We conduct sensitivity analysis over the control variables to include, the inclusion of inverse temperature weights, and outliers.

To summarize, we (1) combine the HS and NM data; (2) assesses whether they meet the selected NM selection criteria; (3) adopt NM's quality weights and infer the appropriate quality weights for HS estimates based on NM's "subjective" criteria; (4) set weights equal to zero for duplicate estimates and cap individual study weights at one; (5) address heteroskedasticity by also weighting by inverse temperature; (6) introduce methodological controls, including compensating surplus and productivity control variables to the preferred HS specification; and (7) cluster standard errors at the model level.

Our preferred estimate implies a GDP loss of 6.7 to 8.3 percent for a 3°C increase, depending on the treatment of catastrophic risks; see Specification (1) in Table 4. Because the DICE damage function should include an insurance premium (Nordhaus 2019), the appropriate estimate for calibrating the DICE damage function should be on the higher end, close to 8.3 percent. Accounting for productivity effects further increases noncatastrophic and total damages to 8.8 to 10.4 percent of GDP, respectively, since productivity effects are three times larger than the direct market effects of climate change. Despite leaning into the NM selection criteria and weights, we find that the resulting damage estimates are very similar to HS.

These preferred results are relatively robust to data selection (treatment of duplicate and compensating surplus estimates, definition of global damages), data entry (definition of catastrophic and quality), and methodological assumptions (inclusion of control variables and the use of the fixed effects model), particularly total damages that combine the noncatastrophic impacts and catastrophic risks of climate change; see Table SM.8 in the Supplementary Material. Specifically, despite some variation by assumptions (most notably for data entry), we find that the estimates always exceed the NM estimates, sometimes reaching or exceeding the HS noncatastrophic damage estimates of 6.7 to 8.0 percent and total damages of 9.0 to 10.3 percent for a 3°C increase, based on the specific modeling assumptions. As theorized, we note that dropping compensating surplus estimates has *no* effect on the results because of the inclusion of a compensating surplus control variable, such that we drop climate damage estimates from Maddison (2003) and Maddison and Rehdanz (2011) in the following section. For an in-depth discussion of the sensitivity analysis, see Section SM.4 of the Supplementary Material.

## 5. Addressing Subjectivity in Weights

Thus far, we have demonstrated that objective and subjective decisions drive the differences in meta-regression results between HS and NM. Moreover, if we combine the standard meta-regression techniques laid out in HS, including controlling for methodological variables and using a fixed effects model, with the subjective assumptions in NM, including selection criteria and quality weights, we find that our preferred estimate matches HS quite closely. To address subjectivity in the selection criteria and data entry, we conducted sensitivity analysis to these assumptions, finding that total climate damages are relatively robust, though the range of outcomes is wide, with some estimates above and others below the HS estimate. However, all estimates exceed NM. Up to this point, we simply adopted NM's quality weights and extrapolated them to the HS estimates (not in NM) using Table 1 as a guide, despite the clear subjective judgment in the NM specification.

To limit the subjective nature of weights, this section develops a series of weighting rules for the climate damage context. We then apply our weights using two methodologies: (1) the quality effects estimator (Doi et al. 2015), and (2) the fixed effects estimator, where we explicitly control for the various weight components. *A priori*, updating the weights can shift the final estimate in any direction.<sup>31</sup> Using our preferred weighting rules and estimator, we find that quality weights have little effect on the coefficient corresponding to  $t_2$  (in other words, direct noncatastrophic climate impacts), though the effect on productivity and catastrophic impacts is more uncertain.

### 5.1 Developing Objective Components of Weights

To reduce subjective bias, it is important to consider all the properties that estimate weights should have *a priori*. One way to achieve this objectivity is to develop a set of weighting rules ( $\omega_{i,rule}$ ) and determine which components of these rules require subjective judgment. With weighting rules and transparency in the underlying decisions of those weights, it is easier for analysts and reviewers to explore the reasonableness of the underlying assumptions and to conduct sensitivity analyses. Below, we discuss several weighting rules that have been implicitly or explicitly applied in the climate impact meta-analysis literature. We also develop several additional weighting rules.

First, because of the common issue of heteroskedasticity, it is important to weight estimates based on their level of precision. In the case of statistical and survey

---

<sup>31</sup> On the one hand, if we followed the NM's implicit weighting of estimation methodologies to its logical conclusion of focusing solely on enumerative estimates, we would produce a damage estimate lower than HS and more similar to NM. Other the other hand, we could place weights solely on the science-based and survey methods, producing an even higher estimate.

estimates, analysts ideally use the reported standard errors or the number of observations as measurements of precision (Nelson and Kennedy 2009; Stanley et al. 2013). However, in the case of global climate damage estimates, these statistics are often unavailable. Instead, HS applies inverse temperature as a proxy variable for standard errors. Implicitly, HS assumes that the variance of a damage estimate is linear in temperatures up to 4°C and then infinite thereafter, such that the weight drops to 0 percent. An alternative approach is to leverage the limited number of standard deviation estimates to estimate the appropriate functional form (Tol 2009). Taking this empirical approach moving forward, we estimate that damage variance is a quadratic function of temperature ( $\hat{\sigma}_t^2 = 6.8 * t_i^2$ ) instead of a linear function, based on statistical fit and precedence (Tol 2009). Using this relationship, we then extrapolate damage variances for all damage estimates to use in our preferred specification for weights ( $\omega_{i,heteroskedasticity} = \frac{1}{\hat{\sigma}_t^2}$ ). We conduct sensitivity analysis and confirm that damage variance increases rapidly above 3.5°C nonlinearity (see Figure 2) and that a quartic function also performs well; see Section SM.5 in the Supplementary Material for further discussion.

Second, it seems logical that estimates should be down-weighted based on their age as well as on temperature. As time passes, the climate and scenario information underlying a study becomes outdated. However, the appropriate functional form for the age-specific weight  $\omega_{i,age} = g(Age_i)$  is unclear because data are not available, as they are for heteroskedasticity. Therefore, judgment with respect to the functional form is necessary. An alternative is to include years since publication as a control variable in the regression analyses (Howard and Sterner 2017).

Third, multiple estimates per study weights ( $\omega_{i,study}$ ) should sum to one within a study before down-weighting for study characteristics like temperature and age (Nordhaus and Moffat 2017; Howard and Sterner 2017, appendix). In other words,  $\sum_{i \in S} \omega_{i,S} = 1$  for all estimates  $i$  in study  $s$ . With a small sample size, authors should select one estimate per study unless the estimates are truly independent (i.e., they should not include virtually the same estimate extrapolated to multiple temperature levels) (Howard and Sterner 2017). Subjectivity enters the processes when the analyst attempts to determine the relative weights of estimates within a study. If the estimates are truly independent, then equal weighting is likely the least biased approach (e.g., Nordhaus 1994b; Howard and Sylvan 2015). This approach is less satisfactory when one estimate has stronger theoretical and/or empirical grounding than the other (e.g., Cline 1992; Nordhaus 2014; Weitzman 2012). If an inverse temperature weight is also applied by the analyst, care should be taken to avoid double-penalizing the more speculative estimate if its speculative nature is partially the result of its higher temperature (e.g., Cline's 1992 damage estimate at 10°C).

Fourth, like older studies, papers that are duplicate estimates should be down-weighted to account for their lower informational content (i.e., assigned a weight  $\omega_{i,duplication}$ ). As noted in HS, updated estimates by an author using the identical estimation strategy employed in an earlier study should be interpreted as an update, not an independent estimate. Similarly, papers that adopt the damages of another

study are derivative estimates and provide little to no new information. However, these weighting issues become even more complicated because of the overlapping nature of most climate research in the economic field. Enumerative damage estimates are often used to calibrate CGE models, but the latter estimates can capture productivity effect excluded from the former (see previous discussion of the independence of the *productivity* variable). Clearly, the informational content is not full, but the exact value is unclear. Similarly, how much additional information is provided by studies using the same underlying data (e.g., Ng and Zhao 2011 and Nordhaus 2006, 2008b, all of which use the G-ECON data set) or models (e.g., GTAP is the basis of virtually all CGE estimates, and Hanemann 2008 modifies Nordhaus and Boyer 2000). Analysts' opinion on information overlap is required to determine the weight of these related estimates, introducing subjectivity into this process; see Table 1 for a comparison of the HS and NM approaches.

Fifth, a paper should be down-weighted for having a poor identification strategy (i.e., assigned a weight  $\omega_{i,quality}$ ). Although there can be some subjectivity in the determination of papers that have poor identification strategies—an analyst may prefer experiments to instrumental variables or have a strong opinion on the quality of weather-based damage estimates—it is generally clear within an estimation method which strategies are stronger. This type of weighting applies most readily to statistical studies where the strength of identification strategies may differ across studies. Even so, it is unclear how much information is provided by a poor estimate relative to a good estimate (i.e., the penalty function is unknown), such that the selection of weight introduces additional subjectivity. In the previous section, we assigned a weight of 10 percent following NM. Critically, a determination of low quality should be independent of other weight components (e.g., age and temperature) and methodological controls. If an estimate's poor identification strategy is partially or fully addressed by the application of methodological controls, such as the inclusion of the indicator variable *cross* in HS to control for the common problem of omitted variable bias in cross-sectional studies, then analysts should no longer include it as a control variable or should no longer consider this issue as part of the weight.

Last, analysts should weight estimation methodologies to reflect their relative information content. Previous meta-analyses of climate damages implicitly weight estimation methodologies to reflect the combined effect of nonrandom selection of estimation methods and the assignment of methodological quality (via estimate-specific quality weights). In the previous section, the enumerative method is strongly favored because it is the estimation method used in approximately 37 percent of damage estimates. To avoid the implicit weighting of methodologies, we separate the weighting of estimates within an estimation strategy (i.e., the above weight components) from the weights given to estimation strategies overall. We then split the methodological (or estimation strategy) weight into two components: (1) accounting for the number of studies  $s$  applying a particular estimation methodology  $m$  to account for the overlapping nature of these estimates ( $\omega_{i,s,m}$ ); and (2) accounting for the subjective quality of the estimation method  $m$  ( $\omega_{i,m,quality}$ ). The former part of the methodological weight  $\omega_{i,s,m}$  is necessary because the marginal information

content of an additional study depends on its estimation methodologies' proportional representation in the data set, while the latter component  $\omega_{i,m,quality}$  assesses the overall quality of the methodology for providing information. To be consistent with our weighting of multiple estimates per study, such that the study weights sum to one, the weights within methodologies should sum to one.

The final objective weight is

$$(4) \quad w_i = \prod_{rule\ r} \omega_{i,r} = [\omega_{i,m,quality} \times \omega_{i,s,m}] [\omega_{i,quality} \times \omega_{i,duplication} \times g(Age_i) \times \omega_{i,s}] \frac{1}{\hat{\sigma}_i^2}$$

subject to  $\sum_{s \in m} \omega_{i,s,m} = 1$  and  $\sum_{i \in s} \omega_{i,s} = 1$ .

Note that the second bracketed term is consistent with the weight specification in NM, which we will refer to as the estimate-specific component of the weight, and  $\frac{1}{\hat{\sigma}_i^2}$  is consistent with the fixed effects estimator in HS, such that we are introducing the methodology-specific weight only in the first bracket. For some estimators, the variance weight is separated from the quality weight, such that the final objective weight is  $w_i \hat{\sigma}_i^2$ .

## 5.2 Subjective Assumptions Underlying Weights

Given the large number of subjective decisions, we specify a base specification and then conduct a sensitivity analysis. Our preferred specification for total estimate weight is

$$(5) \quad w_i = [\omega_{i,m,quality} \times \omega_{i,s,m}] \left[ \omega_{i,quality} \times \omega_{i,duplication} \times \left[ \frac{Year_i - 1993}{2021 - 1993} \right] \times \omega_{i,s} \right] \frac{1}{\hat{\sigma}_i^2} = \frac{\alpha_{i,m} \times \beta_{i,s}}{\hat{\sigma}_i^2}$$

For the estimate  $i$ 's specific part of the weight  $\beta_{i,s}$ , we assume the following:  $\omega_{i,duplication}$  is 1 if an independent estimate and 0 if duplicate;  $\omega_{i,quality}$  is 1 if good quality and 10 percent if low quality, following NM;  $\omega_{i,s} = \frac{1}{n_{i,s}}$  where  $n_{i,s}$  is the number of independent estimates in study  $s$  unless one estimate is clearly methodologically superior, such that it is assigned 90 percent (applies only to scientific methodology, as discussed above); and where the age adjustments are linear, such that every seven years (when a new IPCC document is published) the weight is reduced by 25 percent (and thus the weight equals 0 for studies before 1994).<sup>32</sup> However, given the agnostic

---

<sup>32</sup> As a proxy for quality, estimates should not be weighted based solely on an author's experience or a study's number of citations. As demonstrated in the expert elicitation literature, weighting estimates by these factors does not improve predictive ability (Cooke et

view of weather-based studies in HS, we weight these studies at 50 percent to account for the ongoing debate underlying this methodology. Because cross-sectional studies are classified as low quality, we exclude the cross-sectional control.

For the methodology  $m$ 's component of the weight  $\alpha_{i,m}$ , we make two assumptions. First, we assume equal weighting of methodologies in terms of quality (i.e.,  $\omega_{i,m,quality} = 1 \forall m$ ), based on the principle of insufficient reason, since we lack sufficient data to determine the appropriate methodological weights (Ho et al. 2019). Second, we assume that our weighting of multiple studies per method ( $\omega_{i,s,m}$ ) should be consistent with our weighting of multiple estimates per study. However, because we combine equal weighting and relative weights (for scientific studies) above, we take two approaches:  $\omega_{i,s,m} = \frac{1}{N_{i,s,m}}$  where  $N_{i,s,m}$  is the number of studies  $s$  corresponding to estimation method  $m$  (henceforth equal weighting) or  $\omega_{i,s,m} = \frac{1}{\sum_{s \in m} \beta_{i,s}}$  where  $\beta_{i,s}$  is the estimate-specific component of the weight, such that estimate-specific weights sum to one within a methodology (henceforth proportional weighting).<sup>33</sup>

### 5.3 Quality Effects Estimator

As discussed in Section 2.2, to address the hierarchical structure of meta-analysis, weighted least squares is the workhorse estimator. Specifically, the random effects (RE) estimator applies where the weight equals  $w_i = \frac{1}{\sigma_{i,k}^2 + \tau^2}$  because of its efficiency when  $\tau^2 > 0$  in Expression (2). Despite the efficiency of the fixed effects (FE) estimator ( $w_i = \frac{1}{\sigma_{i,k}^2}$ ) when there are sufficient controls for methodological and population differences (i.e.,  $\tau^2 = 0$ ), RE is often used by default because the test for heterogeneity is inconclusive and FE is inefficient with even small amounts of unexplained between-study variation (Doi and Thalib 2008). However, in a series of papers Dr. Suhail Doi, a clinical epidemiologist, questions the default nature of the RE model with respect to the FE model and more generally. Critically, the RE estimator is unsatisfactory because between-study variation is not random in nature: it arises from uncontrolled methodological or population differences, such that the homogeneous adjustment to the denominator of the RE weights ( $\tau^2$ ) to address a problem of nonrandom heterogeneity is better addressed by estimate-specific adjustments ( $\tau_i^2$ ) (Doi and Thalib 2008, 2009; Stone et al. 2020). In economics, including additional methodological controls and using FE are the standard solution, though analysts are often unable to apply this approach because of an insufficient number of studies (to avoid spurious findings, the rule of thumb is 10 studies per covariate; Gagnier et al. 2012; Higgins et al. 2020), lack of the necessary data, difficulty in defining variables, or unclear pathways of causation (Thompson and Higgins 2002). In our context, this approach is undesirable because we already have fewer than 10 studies per covariate, even after dropping the

---

al. 2008, 2021). This issue is compounded by the existing problem that more established academics already have more published damage estimates often favoring a specific strategy, and this could bias the resulting estimates if duplication is not sufficiently addressed.

<sup>33</sup> This latter method ensures that estimate-specific weights affect only the relative weight of the estimate within its corresponding methodology.

cross-sectional control and excluding compensating surplus estimates and its corresponding control variable.

In the medical sciences, an alternative is to collect information on this nonrandom heterogeneity using quality assessments, which measure the proportion of quality safeguards applied (Valentine and Cooper 2008), to proxy for the probability of an unbiased estimate. However, the quality score estimator—a method that simply uses the quality score as the weight in weighted least squares, as implemented in NM—has received scrutiny because it does not solve for biases and can even reduce estimator efficiency. This is because quality scores do not represent the direction, degree, or even the ordinal ranking of bias and are dependent on the quality scale applied (Greenland 1994; Greenland and O'Rourke 2001; Valentine 2009; Ahn and Becker 2011; da Costa et al. 2013). To address the shortcomings of the simple quality score estimator, Doi developed and refined the quality effects (QE) estimator to replace the RE estimator (Doi and Thalib 2008, 2009; Doi and Barendregt 2013; Doi et al. 2015; Stone et al. 2020). Based on several advantages of the QE, including efficiency, independence from direction or magnitude of bias, and nonreliance on the full information content of the quality score (such that it collapses to the RE estimator if weights are randomly assigned), we adopt it as our preferred estimator; see Section SM.6 in the Supplementary Material for a more in-depth discussion of the QE estimator.

## 5.4 Results

Using the quality effects estimator, we estimate our preferred specification with the combined NM and HS data, the HS selection criteria (i.e., dropping the compensating surplus estimates), and the equal and proportional weighting approaches; see Table 4. The results are nearly identical between equal and proportional weights, with the primary coefficient of interest corresponding to  $t_2$  matching HS quite closely, such that noncatastrophic damages increase from 6.7 percent in HS to 6.9 percent after weights are applied.<sup>34</sup> However, catastrophic and productivity impacts increase relative to HS, such that more expansive definitions of climate damages that include catastrophic and productivity impacts exceed the HS estimate of 10.3 percent of GDP, increasing to 16.4 to 17.3 percent.<sup>35</sup>

As discussed in the previous subsection, an alternative to quality weighting is to include weighting components, specifically estimate quality and study age, as

---

<sup>34</sup> The standard errors also narrow relative to HS, though this is not the result of using QE. Instead, the narrowing results from the combination of including high-temperature damage estimates and redefining the catastrophic indicator variable to exclusively refer to the risk premium concept.

<sup>35</sup> Reestimating the model without a catastrophic risk premium included (see Supplementary Material, Table SM10, Specifications 5 and 6), we find that the results are robust to focusing exclusively on noncatastrophic effects where the ratio of coefficients to data is above or closer to 10 to one.

additional control variables in an extended fixed effects model.<sup>36</sup> Estimating the fixed effects model, we find that including quality and age control variables significantly increases the coefficient corresponding to  $t_2$ , as lower quality estimates bias damages upward and older estimates significantly bias damages downward. Specifically, direct noncatastrophic effects increase to 11.1 percent for a 3°C increase, though productivity and catastrophic impacts decline, such that the more expansive definitions of damages remain unchanged or decline. However, care should be taken in interpreting these final fixed effects results because the ratio of estimates to coefficients is far below the recommended ratio of 10 to one.

For purposes of transparency, we conduct extensive sensitivity analysis for estimate- and methodology-specific weights for all the above estimators; see SM.8 for an in-depth discussion. Overall, our results suggest that our preferred QE estimates are highly robust to alternative subjective assumptions. This is particularly true for the primary coefficient of interest corresponding to variable  $t_2$ . If anything, our preferred estimator may underestimate climate damages, since we find that assuming a zero weight for estimates above 4°C (to account for the explosion of standard errors around this temperature level) and our fixed effects regression, extended to include quality and age variables, increases our coefficient corresponding to  $t_2$ . The coefficients corresponding to the catastrophic and panel variables are more sensitive across specifications, including whether science-based and weather-based estimates are included.

Analysts cannot escape issues of author discretion when determining weights. Authors face a variety of subjective weighting decisions, particularly on the relative weight given to estimates: the relative weight of differing methodologies (enumerative, statistical, etc.) and the functional form to employ in temperature and study age weighting. This is in addition to subjective selection decisions, like where the demarcation between regional and global damage estimate is placed, what qualifies as a duplicate or low-quality estimate, and to what extent an estimate should be down-weighted. Analysts must transparently document their *a priori* decisions and their reasoning on complicated decisions. A future direction of research could employ expert elicitation to determine the consensus or central view in the literature for particular decisions to avoid basing policy recommendations on individual authors (Pindyck 2019; Howard and Sylvan 2015).

Interestingly, we find that many of the subjective decisions have little effect on direct noncatastrophic damages (i.e., the coefficient corresponding to  $t_2$ ), implying damages of approximately 7 percent of GDP for a 3°C increase. Instead, these subjective assumptions tend to affect the coefficients corresponding to the catastrophic ( $cat\_t_2$ ) and productivity ( $prod\_t_2$ ) impacts, such that a wider definition of climate damages is more uncertain. Our preferred specification implies that combined direct and productivity noncatastrophic effects ( $t_2+prod\_t_2$ ) may reach 12.8 percent for a 3°C

---

<sup>36</sup> We reject the random effects model; see Section SM.7 for more discussion.

increase, which increases further to 15.9 percent when a catastrophic risk premium is included.

## 6. Conclusion

Nordhaus (2019) highlighted two contemporary damage functions estimated using meta-analysis with two completely different policy recommendations. In this paper, we show that these differences arise from different regression models and estimators, particularly Nordhaus and Moffat's (2017) choice not to address heteroskedasticity and the partially overlapping effects captured by differing estimation methodologies, and from subjective judgments made at various steps in the data development stages: search, selection, imputation, and weighting. Adapting the latest meta-analysis techniques for the climate damages context is relatively straightforward and indicates that direct damages are between 7 and 10 percent of GDP for a 3°C increase in global average surface temperature, depending on the inclusion of productivity effects and/or a risk premium for catastrophic impacts. However, exploring alternative subjective decisions in the data selection process, including the selection of quality weights, makes an even wider range of damages possible.

To address the subjective element in meta-regression, we transparently develop quality weights and introduce the quality effects estimator from the health literature. Applying this estimator with our weights, we find that direct noncatastrophic damages from climate change are approximately 7 percent of GDP for a 3°C increase. This estimate is highly robust to a variety of subjective decisions, though focusing exclusively on estimates developed using the enumerative methodology still results in a significant decline in direct noncatastrophic damages. In contrast, the coefficients corresponding to productivity effects and the catastrophic risk premium are more sensitive to these subjective decisions, though our preferred specification implies that total climate damages could reach up to 16 percent when we factor in the more uncertain productivity effects and catastrophic risks.

To further improve meta-analyses of climate change damages, future studies need to engage more fully with the subjective decisions that analysts face when combining damage estimates. To improve the search process, future work with data scientists is necessary to refine data searches in the diffuse field of climate damages. Given the small sample size of global estimates, it is critical that researchers balance the identification of relevant studies with imprecise identification (and hence overidentification). Additional work is also necessary to develop weighting rules, potentially drawing on existing work in the medical field. In this paper, we take a first step by making transparent the implicit weighting assumptions in the literature and highlighting which components are subjective in nature; additional desirable features of estimate weights are specified.

Above all, we must all be humble in the face of the truly gigantic difficulties involved in understanding both the science of climate change and the best way of valuing future consequences. When we speak in this study of the size of certain effects, we are reflecting the extent of consensus in the current literature. The actual social changes we may face because of a changing climate are probably not *knowable* in a deeper sense. They will depend on very complex interactions between nature, human behavior, and policy. To some extent, time will help as we gain some knowledge of damages and the costs they impose. For example, the IPCC (2021) reported considerable progress in its latest (sixth) assessment report when it comes to assigning causal effects between climate changes and observed events. Perhaps researchers will find new and better ways to estimate damages, as exemplified by the work of the Climate Impact Lab (Carleton et al. 2020; Rode et al. 2021) and Resources for the Future (Bressler et al. 2021; Newell et al. 2021), matching similar ongoing progress made on discount rates (Li and Pizer 2021; Newell et al. in press) and socioeconomic scenarios (Ho et al. 2019; Rennert et al. 2021). However, we cannot afford to wait: future damages are likely at any date to dwarf current costs if the processes are cumulative or nonlinear or involve thresholds. We have no choice but to both undertake new studies and continue to survey the literature in a fair and objective way.

## 7. References

- Ackerman, F., E. A. Stanton., and R. Bueno. 2012. CRED v. 1.4 Technical Report. Somerville, MA: Stockholm Environment Institute.
- Ahn, S., and B. J. Becker. 2011. Incorporating Quality Scores in Meta-Analysis. *Journal of Educational and Behavioral Statistics* 36(5): 555–85.
- Bluedorn, J. C., A. Valentinyi , and M. Vlassopoulos. 2010. The Long-Lived Effects of Historic Climate on the Wealth of Nations. 2010 Meeting Papers No. 627, Society for Economic Dynamics.
- Bosello, F., & Parrado, R. 2014. Climate Change Impacts and Limited Market-driven Adaptation. Review of Environment, Energy and Economics-Re3.
- Bressler, R. D., F. C. Moore, K. Rennert, and D. Anthoff. 2021. Estimates of Country Level Temperature-Related Mortality Damage Functions. *Scientific Reports* 11(1): 1–10.
- Burke, M., S. M. Hsiang, and E. Miguel. 2015. Global Non-Linear Effect of Temperature on Economic Production. *Nature* 527(7577): 235–39.
- Carleton, T. A., A. Jina, M. T. Delgado, M. Greenstone, T. Houser, S. M. Hsiang, et al. 2020. Valuing the Global Mortality Consequences of Climate Change Accounting for Adaptation Costs and Benefits (No. w27599). Cambridge, MA: National Bureau of Economic Research.
- Cline, W. R. 1992. The Economics of Climate Change. Washington, DC: Institute for International Economics.
- Cooke, R. M., S. ElSaadany, and X. Huang. 2008. On the Performance of Social Network and Likelihood-Based Expert Weighting Schemes. *Reliability Engineering and System Safety* 93(5): 745–56.
- Cooke, R. M., D. Marti, and T. Mazzuchi. 2021. Expert Forecasting with and without Uncertainty Quantification and Weighting: What Do the Data Say? *International Journal of Forecasting* 37(1): 378–87.
- da Costa, B. R., R. Hilfiker, and M. Egger. 2013. PEDro’s bias: Summary Quality Scores Should Not Be Used in Meta-Analysis. *Journal of Clinical Epidemiology* 66(1): 75–77.
- Dell, M., B. F. Jones, and B. A. Olken. 2012. Temperature Shocks and Economic Growth: Evidence from the Last Half Century. *American Economic Journal: Macroeconomics* 4(3): 66–95.
- Dellink, R., E. Lanzi, J. Chateau, F. Bosello, R. Parrado, and K. De BruinK. 2014. Consequences of Climate Change Damages for Economic Growth: A Dynamic Quantitative Assessment. OECD Economics Department Working Papers, No. 1135. Paris: OECD Publishing. Available at <https://doi.org/10.1787/5jz2bxb8kmf3-en>.
- Doi, S. A., and J. J. Barendregt. 2013. *Meta-Analysis I. Methods of Clinical Epidemiology*. Springer Series on Epidemiology and Public Health. Berlin Heidelberg: Springer, 229–52.
- Doi, S. A., and L. Thalib. 2008. A Quality-Effects Model for Meta-Analysis. *Epidemiology* 19(1): 94–100.

- . 2009. An Alternative Quality Adjustor for the Quality Effects Model for Meta-Analysis. *Epidemiology* 20(2): 314.
- Doi, S. A., J. J. Barendregt, S. Khan, L. Thalib, and G. M. Williams. 2015. Advances in the Meta-Analysis of Heterogeneous Clinical Trials II: The Quality Effects Model. *Contemporary Clinical Trials* 45: 123–29.
- Doi, S. A., J. J. Barendregt, and A. A. Onitilo. 2013. Methods for the Bias Adjustment of Meta-Analyses of Published Observational Studies. *Journal of Evaluation in Clinical Practice* 19(4): 653–57.
- Fankhauser, S. 1995. *Valuing Climate Change: The Economics of the Greenhouse*. London: Earthscan.
- Gagnier, J. J., D. Moher, H. Boon, C. Bombardier, and J. Beyene. 2012. An Empirical Study Using Permutation-Based Resampling in Meta-Regression. *Systematic Reviews* 1(1): 1–9.
- Greenland, S. 1994. Quality Scores Are Useless and Potentially Misleading. *American Journal of Epidemiology* 140(3): 300–301.
- Greenland, S., and K. O'Rourke. 2001. On the Bias Produced by Quality Scores in Meta-Analysis, and a Hierarchical View of Proposed Solutions. *Biostatistics* 2(4): 463–71.
- Greenstone, M. 2017. Assessing Approaches to Updating the Social Cost of Carbon. National Academies Fifth Meeting of the Committee on Assessing Approaches to Updating the Social Cost of Carbon, May 5. Available at <http://www.impactlab.org/research/assessing-approaches-to-updating-the-social-cost-of-carbon/>. Accessed August 16, 2019.
- Hambel, C., H. Kraft, and E. Schwartz. 2015. *Optimal Carbon Abatement in a Stochastic Equilibrium Model with Climate Change*. NBER Working Papers 21044. Cambridge, MA; National Bureau of Economic Research. Available at [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2579694](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2579694).
- . 2021. Optimal carbon abatement in a stochastic equilibrium model with climate change. *European Economic Review*, 132, 103642.
- Hanemann, W. M. 2008. What Is the Economic Cost of Climate Change? University of California–Berkeley. [http://repositories.cdlib.org/are\\_ucb/1071/](http://repositories.cdlib.org/are_ucb/1071/).
- Hanemann, M. 2008. What is the Economic Cost of Climate Change? UC Berkeley: Department of Agricultural and Resource Economics, UCB. CUDARE Working Paper No. 1071. [http://repositories.cdlib.org/are\\_ucb/1071/](http://repositories.cdlib.org/are_ucb/1071/). Accessed January 9, 2014.
- Harbord, R. M., and J. P. Higgins. 2008. Meta-Regression in Stata. *Stata Journal* 8(4): 493–519.
- Higgins, J. P., J. A. López-López, and A. M. Aloe. 2020. Meta-Regression. In *Handbook of Meta-analysis*, edited by C. H. Schmid, T. Stijnen, and I. White. Boca Raton, FL: CRC Press, 129–50.
- Ho, E., D. V. Budescu, V. Bosetti, D. P. van Vuuren, and K. Keller. 2019. Not All Carbon Dioxide Emission Scenarios Are Equally Likely: A Subjective Expert Assessment. *Climatic Change* 155(4): 545–61.
- Holman, B. 2019. In defense of meta-analysis. *Synthese* 196(8): 3189–211.

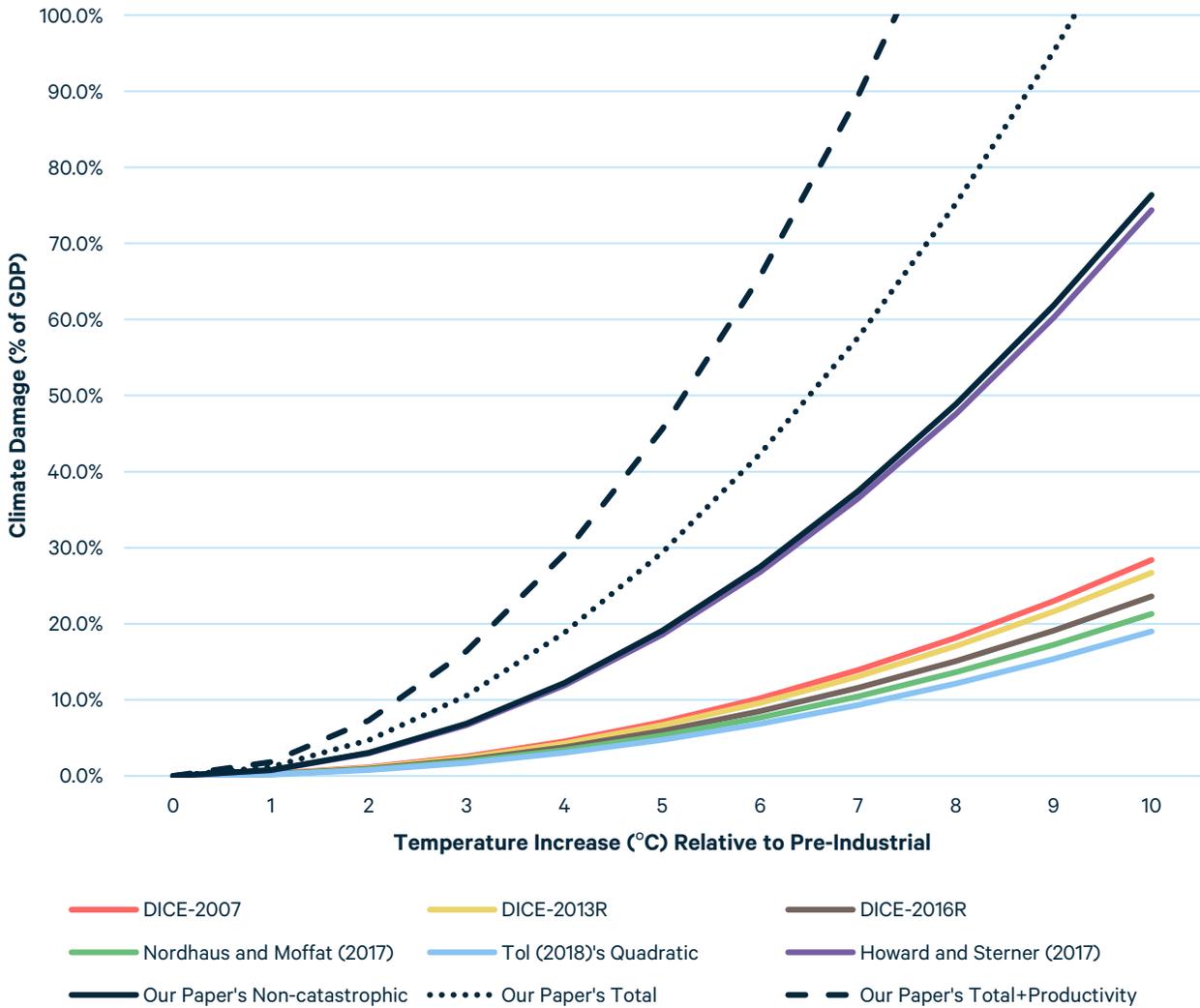
- Horowitz, J. K. 2009. The Income–Temperature Relationship in a Cross-Section of Countries and Its Implications for Predicting the Effects of Global Warming. *Environmental and Resource Economics* 44(4): 475–93.
- Howard, P. H., and T. Sterner. 2017. Few and Not So Far Between: A Meta-Analysis of Climate Damage Estimates. *Environmental and Resource Economics* 68(1): 197–225.
- Howard, P., and D. Sylvan. 2015. Expert Consensus on the Economics of Climate Change. New York: Institute for Policy Integrity, NYU School of Law.
- Howard, P. H., and D. Sylvan. 2020. Wisdom of the Experts: Using Survey Responses to Address Positive and Normative Uncertainties in Climate-Economic Models. *Climatic Change* 162(2): 213–32.
- Holman, B. 2018. In Defense of Meta-Analysis. *Synthese* 196: 3189–211.
- Intergovernmental Panel on Climate Change (IPCC). 2021. Climate change widespread, rapid, and intensifying—IPCC. Geneva, Switzerland: Intergovernmental Panel on Climate Change. <https://www.ipcc.ch/2021/08/09/ar6-wg1-20210809-pr/>. Accessed June 2, 2022.
- Journal of Economic Perspectives (JEP). 2015. Editorial Note: Correction to Richard S. Tol’s “The Economic Effects of Climate Change.” *Journal of Economic Perspectives* 29(1): 217–20.
- Kemfert, C. 2001. Economy-Energy-Climate Interaction: The Model Wiagem. Available at SSRN: <https://ssrn.com/abstract=286835>.
- Kemfert, C. 2002. An Integrated Assessment Model of Economy-Energy-Climate: The Model Wiagem. *Integrated Assessment* 3(4): 281–98.
- Li, Q., and W. A. Pizer. 2021. Use of the Consumption Discount Rate for Public Policy over the Distant Future. *Journal of Environmental Economics and Management* 107: 102428.
- Maddison, D. 2003. The Amenity Value of the Climate: The Household Production Function Approach. *Resource and Energy Economics* 25(2): 155–75
- Maddison, D., and K. Rehdanz. 2011. The Impact of Climate on Life Satisfaction. *Ecological Economics* 70(12): 2437–45.
- Manne, A., R. Mendelsohn, and R. Richels. 1995. MERGE: A Model for Evaluating Regional and Global Effects of GHG Reduction Policies. *Energy Policy* 23(1): 17–34.
- Meyer, A., and T. Cooper. 1995. A Recalculation of the Social Costs of Climate Change. *Ecologist*, Occasional Paper.
- Moore, F. C., U. Baldos, T. Hertel, and D. Diaz. 2017. New Science of Climate Change Impacts on Agriculture Implies Higher Social Cost of Carbon. *Nature Communications* 8(1): 1607.
- Nelson, J. P., and P. E. Kennedy. 2009. The Use (and Abuse) of Meta-Analysis in Environmental and Natural Resource Economics: An Assessment. *Environmental and Resource Economics* 42(3): 345–77.
- Newell, R. G., W. A. Pizer, and B. C. Prest. In press. A Discounting Rule for the Social Cost of Carbon. *Journal of the Association of Environment and Resource Economists*. Advance online publication. <https://doi.org/10.1086/718145>.

- Newell, R. G., B. C. Prest, and S. E. Sexton. 2021. The GDP-Temperature Relationship: Implications for Climate Change Damages. *Journal of Environmental Economics and Management* 108: 102445.
- Ng, P., and X. Zhao. 2011. No Matter How It Is Measured, Income Declines with Global Warming. *Ecological Economics* 70(5): 963–70.
- Nordhaus, W. D. 1994a. *Managing the Global Commons: The Economics of Climate Change*. Cambridge, MA: MIT Press. T
- . 1994b. Expert Opinion on Climate Change. *American Scientist* 82(1): 45–51.
- . 2006. Geography and Macroeconomics: New Data and New Findings. *Proceedings of the National Academy of Sciences of the United States of America* 103(10): 3510–17.
- . 2008a. *A Question of Balance: Weighing the Options on Global Warming Policies*. New Haven, CT: Yale University Press.
- . 2008b. New Metrics for Environmental Economics: Gridded Economic Data. *Integrated Assessment Journal* 8(1).
- . 2010. Economic Aspects of Global Warming in a Post-Copenhagen Environment. *Proceedings of the National Academy of Sciences* 107(26): 11721–26.
- . 2014. Estimates of the Social Cost of Carbon: Concepts and Results from the DICE-2013R Model and Alternative Approaches. *Journal of the Association of Environmental and Resource Economists* 1(1/2): 273–312.
- . 2019. Climate Change: The Ultimate Challenge for Economics. *American Economic Review* 109(6): 1991–2014.
- Nordhaus, W. D., and J. Boyer. 2000. *Warning the World: Economic Models of Global Warming*. Cambridge, MA: MIT Press.
- Nordhaus, W. D., and A. Moffat. 2017. *A Survey of Global Impacts of Climate Change: Replication, Survey Methods, and a Statistical Analysis* (No. w23646). Cambridge, MA: National Bureau of Economic Research.
- Nordhaus, W., and P. Sztorc. 2013. *DICE 2013R: Introduction and User's Manual*.
- Pindyck, R. S. 2019. The Social Cost of Carbon Revisited. *Journal of Environmental Economics and Management* 94: 140–60.
- Rehdanz, K., and D. Maddison. 2005. Climate and Happiness. *Ecological Economics* 52(1): 111–25.
- Rennert, K., B. C. Prest, W. A. Pizer, R. G. Newell, D. Anthoff, C. Kingdon, et al. 2021. The Social Cost of Carbon: Advances in Long-Term Probabilistic Projections of Population, GDP, Emissions, and Discount Rates. *Brookings Papers on Economic Activity*. Available at <https://www.brookings.edu/bpea-articles/the-social-cost-of-carbon/>.
- Rode, A., T. Carleton, M. Delgado, M. Greenstone, T. Houser, S. Hsiang, et al. 2021. Estimating a Social Cost of Carbon for Global Energy Consumption. *Nature* 598(7880): 308–14.
- Schauer, M. J. 1995. Estimation of the Greenhouse Gas Externality with Uncertainty. *Environmental and Resource Economics* 5(1): 71–82.

- Sherwood, S. C., and M. Huber. 2010. An Adaptability Limit to Climate Change Due to Heat Stress. *Proceedings of the National Academy of Sciences* 107(21): 9552–55.
- Stanley, T. D., H. Doucouliagos, M. Giles, et al. 2013. Meta-Analysis of Economics Research Reporting Guidelines. *Journal of Economic Surveys* 27(2): 390–94.
- Stegenga, J. 2011. Is Meta-Analysis the Platinum Standard of Evidence? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 42(4): 497–507.
- Stone, J. C., K. Glass, Z. Munn, P. Tugwell, and S. A. Doi. 2020. Comparison of Bias Adjustment Methods in Meta-Analysis Suggests That Quality Effects Modeling May Have Less Limitations than Other Approaches. *Journal of Clinical Epidemiology* 117: 36–45.
- Thompson, S. G., and J. P. Higgins. 2002. How Should Meta-Regression Analyses Be Undertaken and Interpreted? *Statistics in Medicine* 21(11): 1559–73.
- Tol, R. S. 2009. The Economic Effects of Climate Change. *Journal of Economic Perspectives* 23(2): 29–51.
- Tol, R. S. J. 2013. The Economic Impact of Climate Change in the 20th and 21st Centuries. *Climatic Change* 117(4): 795–808.
- Tol, R. S. 2014. Correction and Update: The Economic Effects of Climate Change. *Journal of Economic Perspectives* 28(2): 221–26.
- Tol, R. S. 2018. The Economic Impacts of Climate Change. *Review of Environmental Economics and Policy* 12(1): 4–25.
- U.S. Environmental Protection Agency (EPA). 2006. Report of the EPA Working Group on VSL Meta-Analysis, Report EE-0494. National Center for Environmental Economics, Washington, DC. [http://yosemite.epa.gov/ee/epa/erm.nsf/vwAN/EE-0494-01.pdf/\\$file/EE-0494-01.pdf](http://yosemite.epa.gov/ee/epa/erm.nsf/vwAN/EE-0494-01.pdf/$file/EE-0494-01.pdf). Accessed December 21, 2016.
- Valentine, J. C. 2009. Judging the Quality of Primary Research. In *The Handbook of Research Synthesis and Meta-Analysis*, edited by H. Cooper, L. V. Hedges, and J. C. Valentine. New York, NY: Russell Sage Foundation, 129–46.
- Valentine, J. C., and H. Cooper. 2008. A Systematic and Transparent Approach for Assessing the Methodological Quality of Intervention Effectiveness Research: The Study Design and Implementation Assessment Device (Study DIAD). *Psychological Methods* 13(2): 130.
- Weitzman, M. L. 2012. GHG Targets as Insurance against Catastrophic Climate Damages. *Journal of Public Economic Theory* 14(2): 221–44.

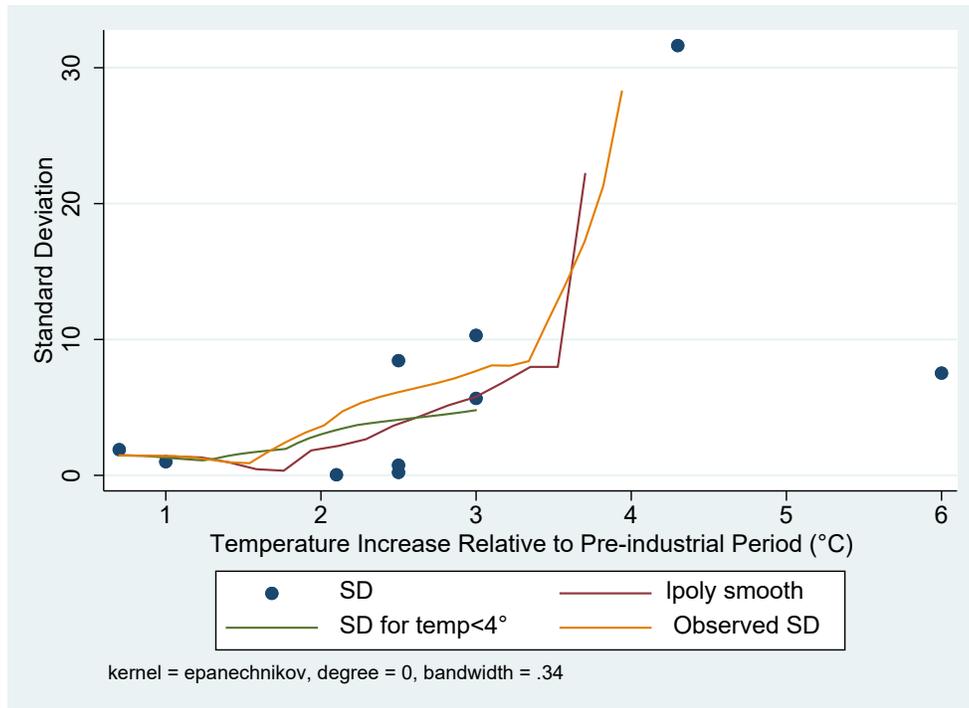
## 8. Tables and Figures

Figure 1. Quadratic Climate Damage Functions in the Literature\*



\* The damage functions from our paper are from our preferred specification in Section 5, assuming equal weighting of studies within methods.

**Figure 2. Kernel-Weighted Local Polynomial Smoothing of Standard Deviation with Respect to Temperature**



**Table 1. Comparison of Weighted Least Squares Methods, NM and HS**

Meta-Analysis Issue	How Do NM Address Issue?	How do HS Address Issue?	Note	Econometric Issue
<b>Data Development</b>				
<b>Search</b>	Combination of formal SRS and informal	Combination of formal and informal	Methods are consistent	Inefficiency and bias (including publication bias)
<b>Selection</b>	Global impacts of climate change	Global willingness to pay to avoid climate change measured as percentage of global GDP	NM include US estimates (Cline 1992) and excludes panel estimates	Duplication bias
<b>Objective Methods</b>				
<b>Dependence</b>	Not Addressed	Clusters standard errors at model level	Only HS address	Biased standard errors
<b>Heteroskedasticity</b>	Not Addressed	Inverse temperature weight as proxy for within-study variation	Only HS address	Inefficiency and biased SE
<b>Methodological Controls</b>	Not Addressed	Control for market-only, catastrophic, and productivity effects	NM also include three compensating surplus estimates that capture only nonmarket effects without sufficient control variables; HS apply general definition of catastrophic	Omitted variable bias
<b>Outliers</b>	Median regression and sensitivity analysis over data and weights	Drop estimates above 4°C and sensitivity analysis over data and estimator	Median regression is robust to outliers in damage space only	Undue influence on results
<b>Subjective Methods</b>				
<b>Quality</b>	10% weight for poor methods	Cross-sectional control and sensitivity analysis	NM implicitly weight time-panel methods at 0%	Inefficiency and potential bias
<b>Age, outdated</b>	0% weight for outdated studies	Test whether estimates vary over time	NM applies to Nordhaus (1994a) and not to similarly aged papers; HS do not control for age despite its statistical significance	Inefficiency and potential bias

**Table 2. Summary of Data, by Data Set, Selection, and Weights**

Variable*	HS: No selection (includes duplicates), no weights	HS: Selection (excludes duplicates), no weights	HS preferred: Selection (excludes duplicates), weights	NM: Selection (includes duplicates), no weights	NM preferred: Selection (includes duplicates), weights	NM: Selection, (excludes duplicates), weights
<b>D_new</b>	6.08	9.45	2.19	1.91	2.03	2.76
<b>T_new</b>	2.72	3.08	1.70	2.53	2.53	2.75
<b>Current</b>	0.51	0.42	0.48	0.33	0.35	0.33
<b>Cat</b>	0.18	0.27	0.15	0.26	0.22	0.13
<b>Gray</b>	0.22	0.19	0.27	0.03	0.01	0.01
<b>Time</b>	12.22	13.38	14.46	12.43	12.21	9.78
<b>Market</b>	0.41	0.42	0.54	0.30	0.34	0.32
<b>Market2</b>	1.00	1.00	1.00	0.89	0.97	0.97
<b>NonMarket2</b>	0.59	0.58	0.46	0.70	0.66	0.68
<b>Productivity</b>	0.22	0.19	0.09	0.20	0.24	0.18
<b>Yale</b>	0.29	0.35	0.25	0.45	0.42	0.41
<b>UCL</b>	0.20	0.12	0.15	0.42	0.48	0.42
<b>Bosello</b>	0.12	0.04	0.03	0.09	0.20	0.13
<b>Nordhaus</b>	0.20	0.23	0.15	0.34	0.39	0.38
<b>Tol</b>	0.06	0.04	0.08	0.16	0.15	0.13
<b>Enumerative</b>	0.37	0.23	0.23	0.51	0.53	0.51
<b>CGE</b>	0.18	0.12	0.07	0.17	0.23	0.17
<b>Science</b>	0.08	0.15	0.05	0.00	0.00	0.00
<b>Statistics</b>	0.27	0.31	0.47	0.26	0.16	0.19
<b>Cross</b>	0.14	0.12	0.26	0.16	0.05	0.04
<b>Compensating surplus</b>	0.00	0.00	0.00	0.11	0.03	0.03
<b>Survey</b>	0.10	0.19	0.17	0.05	0.08	0.13

\* See variable definitions in Table SM.2 in the online Supplementary Material.

**Table 3. Replication of Nordhaus and Moffat's (2017) Weighted OLS Regression Using Howard and Sterner (2017) Data, by Modified Assumptions**

Specification	Preferred	Include data above 4°C	Drop methodological controls and scientific estimates	Drop panel estimates	Drop gray literature	Drop temperature adjustments and clustering	Add missing data and drop HS weights	Add NM weights and change Nordhaus (1994b) and Tol (2013) estimates to reflect NM	Unweighted version of (7)
	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES <sup>a</sup>	D_new	D_new	D_new	D_new	D_new	D_new	D_new	D_new	D_new
T2_new						0.230*** (0.068)	0.268*** (0.076)	0.167*** (0.038)	0.113** (0.039)
t2	0.595** (0.190)	0.318** (0.102)	0.343*** (0.108)	0.249*** (0.059)	0.211*** (0.036)				
mkt_t2	-0.622** (0.226)	-0.345* (0.156)							
cat_t2	0.260 (0.267)	0.362*** (0.103)							
prod_t2	0.113 (0.125)	0.398 (0.237)							
cross	1.700*** (0.343)	1.700*** (0.331)							
Observations	21	26	21	18	16	16	18	18	18
R2	0.722	0.869	0.448	0.406	0.429	0.433	0.423	0.536	0.328
Adjusted R2	0.635	0.837	0.42	0.371	0.39	0.395	0.389	0.508	0.288
Likelihood	-44.36	-72.41	-54.3	-41.59	-34.41	-34.36	-49	-38.51	-45.2
F-statistic	21.95	776.2	10.14	17.98	34.05	11.44	12.46	19.62	8.293
Prob>F	0.000	0.000	0.010	0.002	0.000	0.004	0.003	0.000	0.010
Damage estimate - error relative to Nordhaus and Moffat (2017)									
Estimate at 3°C	6.69%	3.58%	3.86%	2.80%	2.37%	2.59%	3.02%	1.88%	1.27%
Estimate at 6°C	26.78%	14.31%	15.44%	11.21%	9.50%	10.35%	12.06%	7.52%	5.09%
Error	249%	86%	101%	46%	24%	35%	57%	-2%	11%

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

See variable definitions in Table SM.2 in the online Supplementary Material.

**Table 4. Quality Effects and Extended Fixed Effects Regressions Using Combined NM and HS Data, by Methodological Assumption**

Specification	Howard and Sterner (2017)	Section 4, preferred	Section 5, estimated variance only and no weights	Section 5, Preferred		Section 5, alternative estimator using FE with additional controls
				Equal weighting of studies within method	Proportional weighting of studies within method	
Specification	(0)	(1)	(2)	(3)	(4)	(5)
Data <sup>a</sup>	HS Only	HS and NM Combined				
Estimator	FE	FE	FE	QE	QE	FE <sup>c</sup>
VARIABLES <sup>b</sup>	D_new	D_new	D_new	D_new	D_new	D_new
t2	0.595** (0.19)	0.594*** (0.033)	0.689*** (0.091)	0.611*** (0.007)	0.611*** (0.007)	0.984*** (0.132)
mkt_t2	-0.622** (0.226)	-0.536*** (0.073)	-1.076 (0.641)	-0.512*** (0.108)	-0.493*** (0.132)	-0.830 (0.604)
cat_t2	0.26 (0.267)	0.180 (0.234)	0.010 (0.135)	0.412** (0.149)	0.345* (0.188)	-0.072 (0.094)
prod_t2	0.113 (0.125)	0.187 (0.148)	0.844 (0.603)	0.519* (0.293)	0.527* (0.298)	0.083 (0.577)
Cross	1.700*** -0.343		2.333*** (0.754)			
cs_t2		0.366*** (0.033)				
qual_t2						0.824 (0.468)
age_t2						-0.685** (0.299)
Observations	21	29	27	27	27	27
Likelihood	-44.36	-74.55	-55.56	-78.62	-76.67	-55.78
R2	0.722	0.93	0.647	0.989	0.982	0.641
Adjusted R-squared	0.635	0.916	0.567	0.987	0.978	0.539
Weight						
Proxy for damage variance	X	X				
Estimate of damage variance <sup>d</sup>			X	X	X	X
NM's quality weights		X				
Quality weights in Exp. (4)				X	X	
Cluster robust SE <sup>e</sup>	X	X	X			X
Huber-White SE				X	X	

Direct Damages at 3°C Increase to Market and Non-Market Sectors						
Non-catastrophic: t2	6.7%	6.7%	7.8%	6.9%	6.9%	11.1%
Total: t2+cat_t2	9.0%	8.3%	7.8%	10.6%	10.0%	10.4%
Direct Plus Productivity Damages at 3°C Increase to Market Sectors and Non-Market Damages						
Noncatastrophic: t2+prod_t2	8.0%	8.8%	17.2%	12.7%	12.8%	12.0%
Total: t2+prod_t2+cat_t2	10.3%	10.4%	17.3%	16.4%	15.9%	11.4%

Robust standard errors in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

<sup>a</sup> Section 5 only analyzes willingness-to-pay studies dropping compensating estimates.

<sup>b</sup> See variable definitions in Table SM.2 in the online Supplementary Material.

<sup>c</sup> We reformulate quality and age, such that the coefficient corresponding to t2 represents the impact of a current, high-quality estimate.

<sup>d</sup> Standard errors are a linear function in temperature.

<sup>e</sup> Standard errors are clustered at model level.

