RESOURCES *for the* FUTURE

Supplementary Material for Between Two Worlds: Methodological and Subjective Differences in Climate Impact Meta-Analyses

Peter H. Howard and Thomas Sterner

Working Paper 22-11
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.

Acknowledgments

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

Contents

SM 1. Summary of the Tol (2018) Meta-Regression

Using similar data and differing methods, Tol (2018) recently found a nearly identical damage function as DICE-2016R when applying Nordhaus's quadratic damage function. However, we focus in our paper on comparing the results of HS with NM, instead of Tol (2018). Since 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).

Tol does not specify his search methodology, though he does provide selection criteria. Tol (2009) specifies that the analysis will focus exclusively on willingness-to-pay (WTP) estimates and exclude willingness-to-accept (WTA) estimates. He argues that the former, which are lower than the latter, are preferred in the literature. However, Tol (2009, 2014, 2018) does not make this distinction in practice by including compensating surplus estimates: Maddison (2003), Rehdanz and Maddison (2005), and Maddison and Rehdanz (2011).

The primary update that Tol (2018) makes is to expand his data set to include estimates that predate "the first serious study of the global welfare effects of climate change" (Tol 2009) and to include the Nordhaus (2013) meta-regression. Because no systematic review underlies the development of the data, this expansion does not improve his modeling results: many of these new estimates capture only US effects and/or date to the late 1970s and early 1980s. Furthermore, in the case of Nordhaus (2013), the estimate should be excluded for reasons of near-perfect collinearity with the other data points in Tol (2018); it is an earlier meta-analysis using a subset of the Tol (2018) data set, and the estimate is not perfectly collinear because of errors in the underlying data set (Tol 2014; JEP 2015).

From this perspective, the Tol (2018) data set is not an improvement over Tol's earlier data sets (Tol 2009, 2014). The search in Tol (2018) is clearly not systematic, since it ignores omitted studies in HS and NM. Furthermore, Tol (2018) violates Tol's (2009) selection criteria to exclude studies predating 1994 and introduces duplication bias by including the collinear Nordhaus and Sztorc (2013) estimate. These violations also imply a biased estimator. This bias is compounded by Tol (2018), which like the earlier studies, does not apply many of the standard meta-regression techniques laid out in Table 1 of the main text.

SM 2. Applying the HS Selection Criteria to NM Data

Based on the narrower HS selection criteria, none of the studies added through the NM research synthesis (Cline 1992; Nordhaus 2010; Dellink 2012; Kemfert 2001; Hambel et al. 2015) would be included in the HS study because of duplication, geographic extent (e.g., Cline 1992), and lack of necessary data (e.g., Hambel et al 2015). Of the new estimates cited by NM, only Kemfert (2001) could potentially meet the HS selection criteria. Specifically, Kemfert (2001) develops a CGE model calibrated to Tol (2002) and could be seen as a slightly different estimate because of its CGE nature. However, NM incorrectly cite a 2001 working version with several key mistakes, such that the appropriate estimate is −1.8 percent for a 0.25°C increase by 2050; this result is radically different from Tol's (2002) estimates of +2 percent for a 1°C increase, leading Roson and Tol (2006) to question the large magnitude of the indirect effects. For this reason, plus the unrealistic temperature change (inconsistent with business-as-usual scenarios in the literature), we would reject the estimate according to the HS criteria.

The other estimates selected by NM would be more clearly rejected according to the HS selection criteria. Hambel et al. (2015) use a calibration technique (equivalent to an OLS regression using time series data) to estimate negative growth of −0.3 percent for a 1°C increase, but the authors reject their own methodology in an updated 2021 published version of their paper, and the necessary data are unavailable to create a percentage-of-GDP damage estimate because Nordhaus and Moffatt (2017) incorrectly cite a damage that directly affects the growth rate instead of a GDP-level change. Cline (1992) is a US-based estimate and does not meet the global requirement. Nordhaus (2010) and Dellink et al. (2014) are essentially duplicate estimates, with the former applying the DICE-2007 damage function in the context of RICE and the latter relying considerably on the CGE model ICES developed by Bosello (who is also a coauthor in this paper) already included in analysis.

Cline's (1992) US damage estimates should also be excluded in the NM study as well, based on the global selection criterion. Cline (1992) does apply his estimate in a global context, but it is still a US-based estimate. If NM relax their selection criterion to include it, they must consider all US-based estimates, or at least those used by anyone as a proxy for global estimates (e.g., Nordhaus 1991; Titus 1992; Smith 1996), and potentially other country and regional estimates (e.g., OECD and European Union). Interestingly, NM places zero weight on Nordhaus (1994) as outdated, even though this study goes beyond Cline (1992) to a least adjust a US-based damage estimate (i.e., Nordhaus 1991) to represent global effects. If this is outdated, then presumably Cline (1992) is as well.

SM 3. Methodologies for Estimating Climate Damages

As laid out in Howard and Sterner (2017), there are five primary estimation strategies for identifying climate damages: enumerative, expert elicitation, computable general equilibrium (CGE), statistical, and scientific. Starting in 1994 with the first serious empirical research on the economics of climate change, the developers of social-cost integrated assessment models (IAMs) relied heavily on expert opinion (Howard and Sterner 2017). Specifically, modelers calibrated the early damage functions of IAMs using either the enumerative strategy or expert elicitation, as exemplified in DICE-1994 (Nordhaus 1994a). Social-cost IAMs continued to rely on the enumerative strategy to calibrate their damage function, even as new estimation methodologies were introduced in the climate damage context.

The enumerative method is a bottom-up strategy whereby modelers aggregate regional-sector damage estimates from the literature using a combination of author discretion and benefit-cost transfer to fill in the gaps. Given its early adoption, particularly by the early giants in the field and their corresponding social-cost IAMs, this methodology is common; see Table 2 in the main text. Despite its ability to capture direct market, nonmarket, and catastrophic impacts (via either an explicit modeling of tipping points or the inclusion of catastrophic risk premium), it does not capture productivity effects and the resulting general equilibrium effects (hence the development and adoption of CGE-IAMs, as discussed below). Because of its ground-up nature, this methodology frequently omits impacts not explicitly considered by the developer (Howard 2014; Howard and Sylvan 2015, 2020; Howard and Sterner 2017).

Instead of relying on individual authors' discretion, analysts using expert elicitation survey multiple experts on the most likely consequences of climate change. Despite this methodology's early use in the climate economics literature (Nordhaus 1994b; Schauer 1995), the economics literature did not adopt this methodology, perhaps because its perceived subjectivity. However, more recently, expert elicitation has become more popular (Pindyck 2019; Howard and Sylvan 2020, 2021). Despite the development of rigorous empirical methods since then, like the classical method (Cooke et al. 2008, 2021), few analysts have applied these methods in the climate damage context.[1] In expert elicitations about climate change, analysts typically ask experts to consider direct market and nonmarket effects.[2] Providing a specific climate

[1] The most rigorous methodology applied at an aggregate global scale is Howard and Sylvan (2020), who conduct a check for consistency across responses, though recent work by Howard and Sylvan (2021) includes seed questions to check for response accuracy.

[2] To our knowledge, none of the expert elicitations specifically request respondents to consider productivity impacts. However, Howard and Sylvan (2015, 2020) requests asks respondents about the likelihood of growth impacts from climate change finding that most experts on the economics of climate change consider impacts to economic growth likely.

scenario, the analyst asks the expert to provide the 5th, 50th, and 95th percentiles for climate damages or the probability of a catastrophic damage. Thus, the central (i.e., most likely) outcome does not capture catastrophic impacts, though responses to earlier surveys, such as Nordhaus (1994b), were used to calibrate the catastrophic risk premium in the 1999 and 2007 versions of DICE (Howard 2014; Howard and Sterner 2017). Because these elicited values are in terms of percentage of GDP, they are often interpreted as levels, though Howard and Sylvan (2015, 2020, 2021) find evidence that most experts on the economics of climate change believe that climate change will affect economic growth (Howard and Sterner 2017).[3]

After this initial period of reliance on expert opinion, economists began exploring cross-sectional variation using regression analysis to identify climate damages. Economists employed a variety of regression specifications from Ricardian analysis, regressing GDP on temperature, regressing reported measurements of happiness on income and temperature, and a household production function approach. Despite its data-driven approach, however, this methodology was never adopted by the modelers of IAMs to calibrate their damage functions. This is partially because of the well-known problem of omitted variable bias in cross-sectional analysis. By the middle of the first decade of the century, this technique began to wane, and no one identification technique became standard, even as a variety of identification approaches were employed in the literature. Moreover, studies employing statistical methods, including cross-sectional studies, capture only a small portion of noncatastrophic impacts: only market effects in the case of the Ricardian analysis and regressing income on temperature, or only nonmarket effects in the case of studies using the happiness and household production function approaches. Despite this methodology's ability to capture adaptation, it is unable to capture transitional costs. During this period, IAMs continued to use the enumerative approach (Howard and Sterner 2017).

Soon after economists moved away from cross-sectional studies, some economists began to employ the CGE-IAMs to estimate the market damages of climate change. In this approach, modelers typically included climate dynamics and climate impacts in the GTAP model. Specifically, climate impacts were modeled as the loss of inputs in production functions and shifts in the demand for outputs. Often these CGE models calibrated the climate impacts to enumerative-based studies, such that these estimates suffered from many of the shortcomings of this estimation strategy, particularly the reliance on author discretion. Unlike the enumerative approach, in addition to capturing direct market effects, CGE also captures productivity effects

[3] We do not assume that surveys capture productivity impacts, as we are unsure to the extent that experts on the economics of climate change believe climate change will affect growth relative to levels, as the question asks only whether they believe there will be growth impacts in addition to level impacts in Howard and Sylvan (2015, 2020).

and the corresponding effects of climate change on economic growth (Howard and Sterner 2017).[4]

During this time, statistical studies using panel data began to emerge to address omitted variable bias, starting with geographic panel data at a subnational scale and followed by time panel data. Specifically, the panel data structure allows for time fixed effects and/or geographic fixed effects to address the omitted variable bias plaguing cross-sectional studies (Howard and Sterner 2017). Even so, some economists criticize the time panel data approach as capturing weather instead of climate (Nordhaus and Moffat 2017) or relying on untestable assumptions without a theoretical basis (Newell et al. 2021). This is an ongoing area of research, toward which HS and this paper take an agnostic approach. Even so, it should be recognized that the panel approach, particularly the time panel data set, operates though climate's effects on economic growth, though the pathway is unclear. Therefore, the panel methodology captures market effects only potentially via productivity if a time panel is employed (Howard and Sterner 2017).[5]

Finally, science-based damage estimates "use physical thresholds—limits of human physical adaptability to heat stress (Weitzman 2012) and the 2°C limit on global warming recommended by a large body of scientists and international organizations (Nordhaus 2014)—to derive global damage estimates from climate change" (Howard and Sterner 2017). As discussed in the main text, the Weitzman (2012) estimates captures the impacts of catastrophically high temperatures of 6°C and 12°C based on the limits of the human respiratory system, though it does not reflect a catastrophic risk premium, traditionally used in the DICE model. Instead, the Nordhaus (2014) model captures the catastrophic risk premium that scientists implicitly assume based on their belief that a 2°C limit for the increase in global average surface temperature is necessary to avoid catastrophic outcomes. Thus, the science methodology captures the damages associated with the market and nonmarket effects of climate change, though only some estimates capture the catastrophic risk premium.

See Figure SM.1 for a diagram of the types of impacts captured by each methodology. Critically, estimation methodology is strongly related to what types of climate impacts are captured in the resulting damage estimate. Therefore, it is critical to include methodological controls in the meta-regression of damage estimates, since the methodologies are not capturing the same damage concept (i.e., are not pulling from the same distribution of damages). In fact, some methodologies are highly correlated with certain impacts. For example, CGE-IAMs and time panel regression models are

[4] By definition, CGE-IAMs capture general-equilibrium impacts, though transition costs are only sometimes included in the model (Bosello and Parrado 2014).

[5] However, the estimation strategy does not capture long-run adaptation (Howard and Sterner 2017).

the sole methodologies that capture productivity effects.[6] However, they also yield direct market effects, such that the coefficient corresponding to market-only effects is affected when productivity is included. Even after controlling for omitted impacts, care should be taken in interpreting regression results because methodologies are not randomly drawn, such that their distribution in the literature affects the results to some extent.

Please see Howard and Sterner (2017) for a more in-depth discussion of these estimation methodologies.

[6] As noted earlier, surveys may capture productivity impacts to some extent, though it is unclear.

SM 4. Sensitivity Analysis for Section 4 to Address Methodological and Data Discrepancies

In Section 4 of the main text, we focus on selecting the appropriate data search, selection, and entry discrepancies and methodological assumptions from the HS and NM papers. Specifically, we (1) combine the HS and NM data; (2) assesses whether they meet the 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 (we drop the cross-sectional control because quality is addressed in the quality weight); 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, and increases to 8.8 to 10.4 percent of GDP with the inclusion of productivity effects.

Starting with the preferred model, we find that the results are relatively robust to data selection assumptions, particularly total damages; see Table SM.8. Down-weighing the NM duplicate estimates instead of dropping them has a limited effect on the results, though NM did not appear to identify all earlier and superseded versions of the included models according to HS. Similarly, a stricter definition of global damages has a limited effect, though weaking the definition (i.e., loosening when the cutoff between domestic and global damage estimates occurs) transfers some of the impact from noncatastrophic to catastrophic because of the inclusion of Cline (1992).[7] Likewise, dropping compensating surplus estimates has *no* influence on the results, such that we drop climate damage estimates from Maddison and Rehdanz (Maddison 2003;

[7] The cutoff between domestic and global impacts initially may seem black and white, though all current damage estimates use some element of benefit transfer. On one side of the spectrum, Cline (1992) simply assumes that global damages equal US damages. On the other side is a formal benefit transfer. Most climate damage estimates are somewhere in-between with enumerative damage estimates falling closer to the benefit-transfer side of the spectrum. Closer to Cline (1992), Nordhaus and Yang (1996) adjusts US-based damage estimates (Nordhaus 1991) to better represent damages globally by accounting for differing regional-sectoral breakdowns in GDP. Closer to the formal benefit transfer (between an enumerative estimate of global damages and Nordhaus and Yang (1996) on the spectrum), Hanemann (2008) adjusts a US-based damage estimate (Nordhaus and Boyer 2000) to account for downward bias, and then rescales the corresponding global damage estimate (Ackerman and Stanton 2012). HS draws the line for inclusion somewhere between Nordhaus and Yang (1996) and Hanemann (2008). The key is to ensure that your reasoning is consistent with your *a priori* select criteria and that you apply this reasoning consistently (i.e., apply the line in the same place for each estimate that you assess).

Rehdanz and Maddison 2005; Maddison and Rehdanz 2011) in the following section.[8] Only when we make the strong and abstract assumption (i.e., with an unknown criterion) in which we do not integrate the HS estimates with the NM estimates (maintaining only the latter data set) do we observe a significant decrease. Specifically, damages decline by one-half to two-thirds, though the impact is still approximately 10 to 175 percent higher than that of the corresponding estimate in NM.[9] Thus, the subjective selection assumptions can affect the results, though our total damage estimates are still significantly higher than the NM estimates under all alternative selection assumptions that we consider.

The data entry assumptions that we explore can significantly affect damage estimates. Specifically, redefining catastrophic in a looser way corresponding to HS, who include catastrophic temperature levels and catastrophic draws from the damage distribution, decreases the temperature relationship relative to the preferred specification (consistent with Section 3.3). However, the effect on total damages is somewhat mitigated by a switch to catastrophic impacts. Alternatively, eliminating quality weights altogether increases damages by increasing productivity effects by 32 percent, following the higher weighting of weather-based growth impact studies.

Projected climate damages, particularly total damages, are relatively robust to methodological modifications. With respect to control variables, the temperature-squared coefficient is relatively stable to the exclusion and inclusion of methodological controls, though excluding catastrophic and productivity effects significantly underestimates the costs of climate change. Critically, failing to include a productivity control leads to the confounding of direct and indirect market effects (also known as levels versus growth effects), which generally leads to the omission of productivity effects due to their limited representation in the underlying studies. Similarly, dropping the compensating surplus control and controlling for statistical studies that identify climate damages solely on cross-sectional variation have negligible effect on the results, given the down-weighting of these estimates using weights. With respect to weighting, we find that dropping heteroskedasticity weights (i.e., no longer employing the fixed effects model) slightly increases nonmarket and catastrophic impacts.

Total damages are relatively robust to outliers. Using the MM-regression to address outliers, we find a significant decline in noncatastrophic damages. However, this effect is mitigated by a shift from noncatastrophic to catastrophic damages, such that the decline in total damages is significantly mitigated. In contrast, we find that dropping damage estimates above 4°C decreases productivity effects when combined with the

[8] Potentially, including the CS estimates sheds light on the ratio of WTP to CS as a 68 percent increase in nonmarket impacts. However, this is somewhat confounded by the unique estimation methods.

[9] In the NM data set without duplicates, Nordhaus (2008a) is the sole estimate to include a catastrophic impact. Therefore, we remove that catastrophic impact from this estimate and then add it back after the regression to calculate total impact.

robust estimator, such that addressing outliers shifts catastrophic impacts to noncatastrophic, leading to a lower damage function only when productivity effects are included.

Despite some variation by assumptions, we find that the estimates always exceed the NM estimates, sometimes reaching or exceeding HS's noncatastrophic damages 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.

SM 5. Estimating the Variance of the Damage Estimates

Because few damage studies report their standard errors, we observe a small subset of standard deviations underlying damage estimates with which we can empirically inform our functional form assumption $\mathrm{Var}(D_i) = f(T_i)$. Specifically, we directly observe seven standard errors and construct three more from lower- and upper-bound damage estimates.[10] Using a kernel-weighting local polynomial smoothing of the standard deviations underlying damage estimates versus temperature (see Figure 2 in the main text), we see a clear positive relationship. Below a temperature increase of approximately 3.5°C, the relationship follows approximately a linear or quadratic relationship, with a slight positive effect from focusing solely on observed standard deviations. However, consistent with HS's concerns, the standard deviation explodes for higher temperatures, driven upward by Burke et al. (2015), one of only two standard-deviation estimates above 4°C. To test for best fit using adjusted R-squared, log-likelihood value, and AIC, we regress our standard errors on various polynomial functions of temperature: square root, linear, and quadratic; see Table SM.9. Although the quadratic function outperforms the linear model in all three measures of in-sample fit, the relative probabilities that the linear and quadratic models minimize estimated information loss (as measured by AIC) is relatively similar (the square-root specification is also still a possibility even if it is consistently outperformed). Moreover, the quadratic function implies too strong of a decline in the weight of damage estimates as temperatures increase, resulting in near 0 percent weights for estimates above 1°C. Additionally, the linear model outperforms the quadratic model on out-of-sample fit.[11] Finally, Tol (2009) assumes that the standard deviation is linear in temperature, in contrast to HS's implicit assumption of square-root in temperature. Therefore, we estimate a linear model for the standard deviation using all data, finding that $\widehat{\sigma_i} = 2.6 * t_i$. Sensitivity analysis over the coefficient and the functional form is necessary, as well as allowing for nonlinearities at 4°C, as assumed by HS and consistent with Figure 2. In the case where we drop estimates above 4°C, we assume that $SD = 1.7 * t$ based on Specification (6) in Table SM.9.

[10] To approximate the studies' underlying standard errors, we assume that these low and high values represent the 98th percentile. Specifically, we observe the damage range and standard errors for Nordhaus (1994b) and Howard and Sylvan (2020) and select the 98th percentile because it minimizes the prediction error for these two studies relative to other percentile assumptions.

[11] The linear model outperforms the quadratic model when we estimate the model using observed standard deviations and then project the standard deviations for the calculated estimates (using the lower and upper damage boundaries).

SM 6. Quality Effects Estimator

To address the hierarchical structure of meta-analysis, weighted least squares is the workhorse estimator. Traditionally, the efficient weight to minimize mean squared error depends on whether all studies measure the same treatment effect (i.e., $\tau^2 = 0$ in expression 2 because of sufficient control of methodological and population differences), such that the effect size is homogeneous, or different treatment effects due to methodological or population differences (i.e., $\tau^2 > 0$ because of unobserved methodological variables), such that the effect size is heterogeneous.[12] In the former case, the fixed effect (FE) model ($w_i = \frac{1}{\sigma_{i,k}^2}$) is efficient, whereas in the latter case, the random effects (RE) estimator ($w_i = \frac{1}{\sigma_{i,k}^2 + \tau^2}$) is efficient (Doi et al. 2011; Doi and Barendregt 2013). Because the test for heterogeneity is inconclusive and FE is inefficient with even small amounts of unexplained between-study variation, RE is often used by default (Doi and Thalib 2008).[13]

Critically, the RE estimator is unsatisfactory because between-study variation is not random in nature, instead arising from uncontrolled methodological or factual differences. Specifically, the RE estimator applies a homogeneous adjustment τ^2 to the denominator of all estimate weights to address a problem of nonrandom heterogeneity (Doi and Thalib 2008, 2009; Stone et al. 2020). In economics, methodological controls are the standard solution, though analysts may be unable to control for all relevant methodological and population issues because the number of studies is insufficient (the rule of thumb is 10 studies per covariate; Gagnier et al. 2012; Higgins et al. 2020), the necessary data are lacking, variables are not well defined, or pathways of causation are unclear (Thompson and Higgins 2002). 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, there is an ongoing debate in the medical literature over how to apply the resulting quality scores in analysis. 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

[12] In addition to methodological differences, the research populations may differ between studies. For example, although all studies in our analysis measure global climate damages, the population can differ because of the assumed climate scenario. Because climate scenarios determine the rate of temperature change, which affects when a temperature increase occurs and thus the time available for adaptation, the global population of these studies may differ. Controlling for the rate of temperature change or the impact year is beyond the scope of this paper, though future research should address this difference between studies. Similarly, as time passes, our understanding of the physical and economic impacts of climate change advances. To address changes in climate predictions over time, analysts can control for study age; see Table 4 in the main text.

[13] Doi and Thalib (2008) question whether random effects should be the default assumption.

the direction, degree, or even the ordinal ranking of bias and are dependent on the quality scale applied (Greenland 1994; Greenland and O'Rourke 2011; Valentine 2009; Ahn and Becker 2011; da Costa et al. 2013).

To address the shortcomings of the simple quality score estimator, Dr. Suhail Doi, a clinical epidemiologist, 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). By correctly modeling the data-generating process as nonrandom, the QE estimator improves efficiency relative to the RE estimator. Specifically, the QE estimator replaces between-study variation (τ^2) with the additional variation from study-specific bias (ϕ_j^2). Because ϕ_j^2 is unobservable, Doi develops a methodology for constructing a synthetic proxy variable, using quality scores if we reinterpret the quality variable as the proportion of between-study variation (τ^2) unattributed to variance stemming from study-specific bias (i.e., $\omega_{i,quality} = \frac{\tau^2}{\tau^2+\phi_j^2}$ and $0 \leq \omega_{i,quality} \leq 1$). The final weight equals $w_i = \frac{q_j}{\sigma_i^2} + \hat{\xi}_j$ where $q_j = \frac{\omega_{i,quality}}{\mathrm{MAX}[\omega_{i,quality}]}$, which adjusts for scale dependence to measure the relative probability that study *i* is unbiased compared with the best estimate, and $\hat{\xi}_j$ (see Doi et al. 2015, Appendix A, for the calculation) further adjusts this ranking to reduce the bias stemming from the increased correlation between weights and effect size under the QE estimator.[14] Doi also adjusts the standard error calculation to account for overdispersion (Doi and Barendregt 2013; Doi et al. 2015).

The Stata command *admetan*, qe() runs the QE estimator. The admetan package in Stata with the qe option does not allow for control variables. However, because QE weights are simply functions of the quality weights and the variance of study estimates, we can estimate our parameter of interest in two steps. First, run the admetan package in Stata with the qe option providing study j's damage estimate, standard errors, and quality weights to obtain the QE weight (w_i), and then (2) run a weighted least squares regression using w_i as the weight. Instead of adjusting the standard errors ex post, following Doi and Barendregt (2013) and Doi et al. (2015), Dr. Doi in correspondence with the authors instead recommends using Huber-White sandwich estimator for the variances, for computational ease.

The QE estimator does not eliminate bias and instead aims to improve efficiency by minimizing mean-squared error like the FE and RE estimators.[15] In fact, the QE estimator collapses to the RE estimator if weights are randomly assigned, as would occur when weights contain little to no information about risk of bias. Thus, the QE estimator does not rely on direction or magnitude of bias, nor does it rely on the full

[14] Note that for purposes of applying the QE estimator, we separate $\hat{\sigma}_i^2$ from the other components of the weight.

[15] All these estimators are making a variance-bias trade-off, whereby they introduce bias to reduce the variance of the estimate. As mean-squared error is a combination of the two, these estimators are optimal in this tradeoff. Hence, they are the efficient estimators depending on whether the assumed data generating process holds.

informational content of quality scores (i.e., there only needs to be some correlation between weights and probability of bias). Based on these advantages, we will run the QE as our preferred estimator, assuming that our weights measure the proportion of between-study variation unattributed to variance stemming from study-specific bias.

We will also conduct sensitivity analysis to the estimator. Specifically, we will reestimate the HS model using FE estimators after testing between the two models and replacing inverse temperature weights with our inverse estimated variance weights; see Section SM.7. Based on the medical literature, the HS fixed effect model represents a clear alternative to using quality weights because it combines two accepted strategies: evidence-based rules over which sensitivity analysis can be conducted[16] and application of methodological control variables (Greenland 1994; Greenland and O'Rourke 2001; Valentine 2009; Stone 2019, 2020). In contrast to the QE estimator, the FE estimator should include only variables that have a clear theoretically specific sign, which does not hold true for study age and study quality.

Beyond the lack of theoretical priors, this latter approach of including additional control variables is less desirable, since both our estimators already exceed the rule of thumb—10 studies per covariate—to avoid spurious findings as currently specified (Gagnier et al. 2012; Higgins et al. 2020). Relative to HS, we drop the cross-sectional control, which is already factored into the quality weight, as discussed above, to improve this ratio. For this purpose, we also exclude compensating surplus estimates, instead focusing on willingness-to-pay studies, consistent with HS; this eliminates a control variable for which there are only a limited number of estimates, and it has minimal effect on the results.

[16] For example, HS drop low-quality studies where cutoffs are determined *a priori*. This is the strategy taken in HS, since exploding uncertainty above 4°C should collapse the fixed and random effects weights to zero.

SM 7. Testing Fixed versus Random Effects Models

As discussed in the previous subsection, an alternative to quality weighting can be to include weighting components—specifically, estimate quality and study age—as additional control variables in a random or fixed effect model. Running the fixed and random effects regressions, we consistently fail to reject the null that $\tau^2 = 0$, which is unsurprising, since we find that the variation across studies due to heterogeneity is small (i.e., $I^2 \approx 0$); the assumption has little effect on coefficient magnitude or standard errors.[17] Estimating the fixed effect model, we find that including quality and age control variables significantly increases the coefficient corresponding to *t2* because lower-quality estimates bias damages upward and older estimates significantly bias damages downward.[18] This leads to an increase in noncatastrophic effects when productivity is included, though the productivity and catastrophic impacts collapse to zero, causing the broader definitions of climate damages to stay constant or decline. However, care should be taken interpreting these final fixed effect results because the ratio of estimates to coefficients is far below the recommended ratio of 10 to one.

[17] We are unable to weight estimates within studies using random effects, such that we estimate the fixed and random effects models shown in Table SM.11 assigning equal weight to each estimate, regardless of the number of estimates per study. In other words, we drop the multiple estimates per study weight ($\frac{1}{n_i}$). Comparing results for random and fixed effects estimates, we see that this assumption has negligible effect. Given this additional shortcoming of the random effects model, we selected the fixed effect model as our alternative specification to compare with the quality effects estimator.

[18] When we specify quality and age for the regression, we reformulate the variables such that temperature squared represents the effect of recent, high-quality estimates.

SM 8. Sensitivity Analysis for Equal Weight Specification of Quality Effects Estimator

In the main text, we run three key estimators:

- quality effects with the weights specified in Expression (5), assuming equal weighting of studies within estimation methodologies;
- quality effects with the weights specified in Expression (5), this time assuming proportional weighting of studies within estimation methodologies; and
- fixed effects with an extended set of controls, including NM's quality weight and study age interacted with temperature squared.

Our preferred specification is the quality effects (QE) estimator, though we do not specify which quality weight is preferred because both produce comparable results and there is no clear preferred specification numerically or theoretically. We run the following sensitivity analysis for each estimator but discuss and present the results only for the QE estimator with equal weighting of studies within estimation methodologies, since the results are relatively similar for proportional weights. Results for the other sensitivity analyses are available from the authors on request.

First, we reestimate our QE and extended FE models with and without a control for productivity, since the latter model violates the rule of thumb of 10 estimates per covariate; see Table SM.10, where Specification (2) is our preferred specification. However, the inclusion of the additional covariate has no effect on the coefficient corresponding to *t2* (i.e., the primary coefficient of interest). The inclusion of productivity in Specification (2) is theoretically correct—it captures productivity effects (not captured in the coefficient corresponding to *t2*)—and is empirically more reasonable because it indicates that the exclusion of nonmarket effects significantly decreases the climate damage estimate. We also estimate our models with and without catastrophic impacts, finding that the choice has negligible effect on our preferred specification.

Next, focusing exclusively on the equal-weight specification of the QE model, we conduct additional sensitivity analysis to our various subjective assumptions about the weights; see Tables SM.11–16. We find that the results are highly robust to the inclusion and magnitude of the estimate-quality weight ($\omega_{i,quality}$), though coefficients corresponding to the catastrophic and productivity variables are sensitive to whether science-based and weather-based estimates, respectively, are weighted for quality. The results are also quite robust to the penalty function for study age ($g(Age_i)$) and the duplication weight ($\omega_{i,duplication}$). Similarly, the results are

generally robust to the exclusion and alternative specification of the multiple-estimates-per-study weight ($\frac{1}{n_i}$) and the exclusion of the multiple-studies-per-methodology weight ($\frac{1}{N_j}$). The exception is the coefficient corresponding to the catastrophic component, which decreases in magnitude and significance in some specifications. We also vary the methodology-specific weight (i.e., $\omega_{i,m,quality}$), finding that the results are generally robust to several alternative specifications unless we focus exclusively on the enumerative-specific estimates.

We conduct two additional sensitivity analyses for the subjective weights. First, we impose a series of step-by-step changes (from left to right) to replicate more closely the implicit weights in NM; see Table SM.17. However, the results closely resemble HS's preferred results except that the catastrophic climate impact collapses to zero. Second, we set minimum quality weights of 10 and 25 percent in turn; see Table SM.18. The potential for weights to decline below the minimum values in NM appears not to be driving the results.

In addition to the quality weights, we also conduct a sensitivity analysis over the relationship between the standard deviation and temperature. Specifically, focusing on our preferred specification of the quality effects estimator, we also conduct a sensitivity analysis for our variance approximation $f(T_i)$; see Table SM.19. Again, we find that the results are robust except for with respect to a strong nonlinearity at 4°C. Specifically, dropping damage estimates above 4°C (to account for the explosion of standard errors around this temperature level) shifts catastrophic and productivity impacts to the primary coefficient of interest *t2*. The direct total climate impacts (*t2+cat_t2*) are relatively robust to the standard error specification, though the productivity impact (*prod_t2*) declines to zero with the exclusion of Burke et al. (2015). The standard errors around the coefficient corresponding to *t2* also widen.

Finally, we conduct several additional sensitivity analyses. First, we reestimate the preferred specification assuming the broader definition of catastrophic impacts used in HS. We find that the primary effect is to reduce the magnitude of the coefficient corresponding to catastrophic damages and partially shift the impact to the coefficient on temperature squared; see Table SM.20. The standard errors around the coefficient corresponding to *t2* also widen. Next, we explore the sensitivity of our results to the exclusion of every estimate and study. Of all the estimates, only the exclusion of Weitzman's (2012) damage estimate for a 12°C increase significantly alters the coefficient corresponding to the primary variable of interest t2. However, its exclusion shifts catastrophic to noncatastrophic impacts, implying significantly higher direct noncatastrophic damages. Furthermore, dropping the Weitzman (2012) study altogether limits this shift in impact, since the 6°C estimate has the opposing, though smaller in magnitude, effect. For catastrophic damages, dropping Nordhaus's (2014) damage estimate corresponding to a strict 2°C limit on temperature increase causes the catastrophic impact to collapse to zero. Unlike the case with dropping Weitzman (2012), this result holds even if the study is dropped altogether. For the component of climate damages corresponding to productivity effects (i.e., the coefficient corresponding to prod_t2), only the exclusion of Nordhaus's (2008b) estimate using

the G-ECON data set, Bosello and Parrado's (2014) estimate using ICES, or Burke et al.'s (2015) weather-based estimate affect its magnitude; see Figures S1 and S2.

Overall, our results suggest that our preferred quality effects estimates is highly robust to alternative subjective assumptions. This is particularly true for the primary coefficient of interest corresponding to variable *t2*. These primary results strengthen the support of the HS estimates. When combined with our extended fixed effects regression, our results imply that if anything, HS may underestimate climate damages.

Like the quality effects estimator, we also conduct a sensitivity analysis over our preferred fixed effects regression with quality and age control variables; these estimates are available from the authors with a subset available in Tables SM.10–11. We find that the primary result using this specification generally holds: direct climate damages (i.e., the coefficient corresponding to *t2*) increase in the fixed effects estimator relative to the quality effects estimator. However, the FE estimates are much less stable than the QE estimates because the former are sensitive to the specification of the age variable $g(Age_i)$, the quality weight $\omega_{i,quality}$, and the temperature penalty function $f(T_i)$.

SM 9. References

Bosello, F., F. Eboli, and R. Pierfederici. 2012. Assessing the Economic Impacts of Climate Change. *Review of Environment, Energy and Economics (Re3)*. **http://re3.feem.it/getpage.aspx?id=4569**. Cited 29 Dec 2014

Dell, M., B. F. Jones, and B. A. Olken. 2008. Climate Change and Economic Growth: Evidence from the Last Half Century. No. w14132. Cambridge, MA: National Bureau of Economic Research.

Gillingham, K., W. D. Nordhaus, D. Anthoff, G. Blanford, V. Bosetti, P. Christensen et al. 2015. Modeling Uncertainty in Climate Change: A Multi-Model Comparison. No. w21637. Cambridge, MA: National Bureau of Economic Research.

Hope, C. 2011. The Social Cost of CO2 from the PAGE09 Model. Economics Discussion Paper 2011-39. SSRN. Available at **http://dx.doi.org/10.2139/ssrn.1973863**.

Howard. P. H. 2014. Omitted Damages: What's Missing from the Social Cost of Carbon. New York, NY: The Cost of Carbon Pollution Project. Available at **https://costofcarbon.org/resources/entry/omitted-damages-whats-missing-from-the-social-cost-of-carbon**.

Howard, P., and D. Sylvan. 2021. Gauging Economic Consensus on Climate Change. New York: Institute for Policy Integrity, NYU School of Law.

Interagency Working Group on Social Cost of Carbon. 2016. Technical Support Document: Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis under Executive Order 12866. Washington, DC.

Manne, A. S., and R. G. Richels. 2005. MERGE: An Integrated Assessment Model for Global Climate Change. In *Energy and Environment,* edited by R. Loulou, J.-P. Waaub, G. Zaccour. New York, NY: Springer, 175–89.

Mendelsohn, R., W. Morrison, M. E. Schlesinger, and N. G. Andronova. 2000a. Country-Specific Market Impacts of Climate Change. *Climatic Change* 45(3-4): 553–69.

Mendelsohn, R., M. Schlesinger, and L. Williams. 2000b. Comparing Impacts across Climate Models. *Integrated Assessment* 1(1): 37–48

Nordhaus, W. D. 1991. To Slow or Not to Slow: The Economics of the Greenhouse Effect. *Economic Journal*, *101*(407), 920-937.

———. 2013. *The Climate Casino: Risk, Uncertainty, and Economics for a Warming World*. New Haven: Yale University Press.

Nordhaus, W. D., and Z. Yang. 1996. A Regional Dynamic General-Equilibrium Model of Alternative Climate Change Strategies. *American Economic Review* 86(4): 741–65.

Roson, R., and D. van der Mensbrugghe. 2012. Climate Change and Economic Growth: Impacts and Interactions. *International Journal of Sustainable Economy* 4(3): 270–85.

Roson, R., and R. S. Tol (2006). An integrated assessment model of economy-energy-climate–the model. Wiagem: a comment. Integrated Assessment Journal, 6(1): 75–82.

Smith, J. B. 1996. Standardized Estimates of Climate Change Damages for the United States. *Climatic Change* 32(3): 313–26.

Stone, J., U. Gurunathan, K. Glass, Z. Munn, P. Tugwell, and S. A. Doi. 2019. Stratification by Quality Induced Selection Bias in a Meta-Analysis of Clinical Trials. *Journal of Clinical Epidemiology* 107: 51–59.

Titus, J. G. 1992. The Costs of Climate Change to the United States. In *Global Climate Change: Implications, Challenges and Mitigation Measures,* edited by S. K. Majumdar, L. S. Kalkstein, B. Yarnal, E. W. Miller, and L. M. Rosenfeld. Pennsylvania: Easton, 384–409.

Tol, R. S. J. 1995. The Damage Costs of Climate Change toward More Comprehensive Calculations. *Environmental and Resource Economics* 5(4): 353–74.

———. 2002. Estimates of the Damage Costs of Climate Change—Part 1: Benchmark Estimates. *Environmental and Resource Economics* 21(1): 47–73.

SM 10. Tables and Figures

Table SM 1. Damage Studies, Estimates, and Weights in Howard and Sterner (2017) and Nordhaus and Moffat (2017)

Study	Howard and Sterner (2017)		Nordhaus and Moffat (2017)		HS Weights in Preferred Specification	NM Weights	Our Weights to Replicate NM[a]	Our Adjusted NM Weights
	Temperature	Damage	Temperature	Damage				
Cline (1992)	-	-	2.5	-1.1	-	0.9	0.9	0.9
Cline (1992)	-	-	10	-6	-	0.1	0.1	0.1
Nordhaus (1994b)	3	-3.6	3	-3.6	0.33	0.5	0.5	0.5
Nordhaus (1994b)[b]	6	-6.7	6	-10.4	0.00	0.5	0.5	0.5
Aubrey Meyer and Tony Cooper (1995)	3	-11.5	-	-	0.33	0	0	0.1
Fankhauser (1995)	2.5	-1.4	2.5	-1.4	0.40	1	1	1
Schauer (1995)	2.5	-5.2	-	-	0.40	-	0	1
Tol (1995)	-	-	2.5	-1.9	-	0.1	-	-
Nordhaus and Yang (1996)	-	-	2.5	-1.7	-	0.1	-	-
Mendelsohn et al. (2000a)	2.5	0.1	2.2	0.03	0.40	0.1	0.2	0.2
Mendelsohn et al. (2000a)	2.5	0	2.2	0.07	0.40	0.1	0.2	0.2
Mendelsohn et al. (2000b)	-	-	2	0.08		0.1	-	-
Mendelsohn et al. (2000b)	-	-	3.5	0.01		0.1	-	-
Nordhaus and Boyer (2000)	-	-	2.5	-1.5		1	-	-
Kemfert (2002)[c]	-	-	0.25	-1.8		0.1	0.1	0.1
Tol (2002)	-	-	1	2.3		0.1	-	
Maddison (2003)	-	-	3,1	-2.22		0.1	0.1	0.1

Manne and Richels (2005)	2.5	-1.9	-	-	0.40	-	0	0.25
Rehdanz and Maddison (2005)	-	-	1.24	-0.32		0.1	0.1	0
Rehdanz and Maddison (2005)	-	-	0.84	-0.32		0.1	0.1	0
Hope (2011)	-	-	4.085	-3.04		0.25	-	-
Nordhaus (2006)	-	-	3	-1.05		1	-	-
Dell et al. (2008)	3.4	-0.3	-	-	0.29	-	0	0.1
Nordhaus (2008a)	2.5	-1.8	3	-2.49	0.40	0.25	2.725	1
Nordhaus (2008b)	2.1	-0.3	-	-	0.48	-	1	1
Horowitz (2009)	0.7	-3.8	-	-	1.43	-	0	0.1
Bluedorn et al. (2010)	0.7	0	-	-	1.43	-	0	0.1
Nordhaus (2010)	-	-	3.4	-2.8		0.25	-	-
Maddison and Rehdanz (2011)	-	-	3.1	17.8		0.1	0.1	0.1
Ng and Zhao (2011)	0.7	-1.6	-	-	1.43	-	0	0.25
Ackerman and Stanton (2012) adjusting Hanemann (2008)	2.5	-4.2	-	-	0.40	-	0	1
Bosello et al. (2012)	-	-				1	-	
Dellink et al. (2012)	-	-				1	-	
Roson and van der Mensbrugghe (2012)	2.3	-1.8	3.1	-2.14	0.43	0.1	0.1	0.1
Roson and van der Mensbrugghe (2012)	4.8	-4.6	5.5	-6.05	0.00	0.1	0.1	0.1
Weitzman (2012) via Ackerman and Stanton (2012)	12	-99	-	-	0.00	-	0	0.9

Weitzman (2012) via Ackerman and Stanton (2012)	6	-50	-	-	0.00	-	0	0.1
Tol (2013)	1	1.4	-	-	1.00	-	1.825	1
Bosello and Parrado (2014)	2.5	-0.9	-	-	0.40	-	2.5	1
Nordhaus (2014)	3	-10.6	-	-	0.33	-	0	0.9
Nordhaus (2014)	3	-4.9	-	-	0.33	-	0	0.1
Burke et al. (2015)	4.3	-23	-	-	-	-	0	0.1
Hambel et al. (2015)[c]	-	-	NA	NA	-	0.1	0.1	0.1
Howard and Sylvan (2015)	3	-10.2	-	-	0.33	-	0	0.5
Howard and Sylvan (2015)	1	0	-	-	1.00	-	0	0.5
FUND (Gillingham et al. 2015)	-	-	2	0.2	-	0.3	-	-
FUND (Gillingham et al. 2015)	-	-	3	-0.17	-	0.4	-	-
FUND (Gillingham et al. 2015)	-	-	4	-0.85	-	0.3	-	-
WITCH (Gillingham et al. 2015)	-	-	2	-1.84	-	0.3	-	-
WITCH (Gillingham et al. 2015)	-	-	3	-3.72	-	0.4	-	-
WITCH (Gillingham et al. 2015)	-	-	4	-6.25	-	0.3	-	-
PAGE09 (Nordhaus and Moffat 2017)	-	-	2	-0.72	-	0.3	-	-
PAGE09 (Nordhaus and Moffat 2017)	-	-	4	-2.9	-	0.4	-	-
PAGE09 (Nordhaus and Moffat 2017)	-	-	6	-6.51	-	0.3	-	-

Sum of Weights					12.3574717	12.25	12.25	18.25

[a] Weight to replicate NM without duplicates: set any weight greater than 1 equal to 1

[b] We used our interpretation of the temperature and damage estimate

[c] We corrected this estimate from NM

Table SM 2. Variable Definitions

Variable	Description
Damage Variables	
D_new	Climate damage as percentage of GDP corrected for this paper
Damage	Noncatastrophic damages (D_new with catastrophic damages removed)
Factual Causes -Temperature Variables	
T_new	Increase in average surface temperature (°C)
T2_new	*T_new* squared
t	Increase in average surface temperature (°C) adjusted for study's base period
t2	*t* squared
Methodological Causes	
Cat	Dummy equal to 1 if estimate includes catastrophic damages
Cross	Dummy equal to 1 if study uses cross-sectional data without country fixed effects
Current	Dummy equal to 1 if study measures temperature change relative to current temperature
Gray	Dummy equal to 1 if study is drawn from gray literature
Market	Dummy equal to 1 if estimate includes only market damages
Productivity	Dummy equal to 1 if estimate captures the impacts of climate change on GDP through economic productivity
Time	Year published minus 1994
Age	2019 minus year published
CS	Dummy equal to 1 if study estimates climate damages in terms of compensating surplus instead of willingness to pay
Quality	Subjective measurement of quality from NM or Expression (5) of this paper
Interaction Variables	
mkt_t	Interaction of *Market* with t
mkt_t2	Interaction of *Market* with t2
cat_t2	Interaction of *cat* with t2
prod_t	Interaction of *prod* with t
prod_t2	Interaction of *prod* with t2
cs_t2	Interaction of CS with t2
qual_t2	Interaction of quality with t2
age_t2	Interaction of Age with t2
Groups	

Yale	Dummy equal to 1 if study author was associated with Yale as defined in Tol (2009)
UCL	Dummy equal to 1 if study author was associated with University College of London as defined in Tol (2009)
Bosello	Dummy equal to 1 if study author was Bosello
Nordhaus	Dummy equal to 1 if study author was Nordhaus
Tol	Dummy equal to 1 if study author was Tol
Methods for Estimating Climate Damages	
Enumerative	Dummy equal to 1 if enumerative methodology used to estimate climate damage estimate
CGE	Dummy equal to 1 if computable general equilibrium model used to estimate climate damage estimate
Science	Dummy equal to 1 if science methodology used to estimate climate damage estimate
Statistics	Dummy equal to 1 if statistical methodology used to estimate climate damage estimate
Survey	Dummy equal to 1 if survey methodology used to estimate climate damage estimate

Table SM 3a. One-by-One Regressions Using HS Damage Estimates $\leq 4°C$

Specification	Preferred	Include data above 4 degrees	Drop methodological controls and scientific estimates	Drop panel estimates	Drop gray literature	Drop temperature adjustments and clustering	Drop HS' inverse temperature weights	Change Nordhaus (1994b) and Tol (2013) estimates to reflect NM	Add NM weights
	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7a)	(7b)
VARIABLES[a]	D_new	D_new	D_new	D_new	D_new	D_new	D_new	D_new	D_new
t2	0.595**	0.318**	0.302**	0.595**	0.409**	0.595**	0.626**	0.624***	0.382**
	(0.190)	(0.102)	(0.106)	(0.193)	(0.130)	(0.190)	(0.198)	(0.177)	(0.139)
mkt_t2	-0.622**	-0.345*		-0.680**	-0.494*	-0.622**	-0.670**	-0.650**	-0.408*
	(0.226)	(0.156)		(0.251)	(0.211)	(0.226)	(0.212)	(0.215)	(0.176)
cat_t2	0.260	0.362***		0.260	0.312**	0.260	0.248	0.232	-0.094
	(0.267)	(0.103)		(0.271)	(0.126)	(0.267)	(0.275)	(0.264)	(0.139)
prod_t2	0.113	0.398		0.226	0.226	0.113	0.120	0.113	0.118
	(0.125)	(0.237)		(0.160)	(0.167)	(0.125)	(0.077)	(0.125)	(0.094)
Cross	1.700***	1.700***		3.133**	3.133**	1.700***	1.358*	1.700***	2.276
	(0.343)	(0.331)		(0.982)	(1.019)	(0.343)	(0.693)	(0.343)	(2.084)
Observations	21	26	18	19	16	21	21	21	11
R2	0.722	0.869	0.328	0.765	0.751	0.722	0.783	0.741	0.653
Adjusted R2	0.635	0.837	0.289	0.682	0.638	0.635	0.715	0.66	0.364
Likelihood	-44.36	-72.41	-40.44		-30.79	-44.36	-46.22	-43.65	-18.5
F-statistic	21.95	776.2	8.096			21.95	13.13	23.61	
Prob>F	0.000177	0	0.0216			0.000177	0.0011	0.000135	0.653

Robust standard errors in parentheses

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

a See variable definitions in Table SM.2.

Table SM.3b. One-by-One Regressions Using All Damage Estimates from HS

Specification	Preferred	Drop methodological controls and scientific estimates	Drop panel estimates	Drop gray literature	Drop temperature adjustments and clustering	Drop HS' inverse temperature weights	Change Nordhaus (1994b) and Tol (2013) estimates to reflect NM	Add NM weights
	(0)	(2)	(3)	(4)	(5)	(6)	(7a)	(7b)
VARIABLES[a]	D_new	D_new	D_new	D_new	D_new	D_new	D_new	D_new
t2	0.318**	0.343***	0.318**	0.244***	0.318**	0.264***	0.397***	0.211***
	(0.102)	(0.108)	(0.104)	(0.005)	(0.102)	(0.071)	(0.092)	(0.021)
mkt_t2	-0.345*		-0.403*	-0.329*	-0.345*	-0.309**	-0.423**	-0.265*
	(0.156)		(0.187)	(0.157)	(0.156)	(0.101)	(0.149)	(0.133)
cat_t2	0.362***		0.362***	0.431***	0.362***	0.380***	0.284**	0.077***
	(0.103)		(0.105)	(0.006)	(0.103)	(0.071)	(0.093)	(0.021)
prod_t2	0.398		0.236	0.511	0.398	0.433*	0.398	0.161
	(0.237)		(0.156)	(0.343)	(0.237)	(0.234)	(0.237)	(0.125)
Cross	1.700***		3.133***	3.133***	1.700***	1.358*	1.700***	3.315
	(0.331)		(0.952)	(0.962)	(0.331)	(0.669)	(0.331)	(2.872)
Observations	26	21	23	21	26	26	26	13
R2	0.869	0.448	0.909	0.892	0.869	0.933	0.874	0.721
Adjusted R2	0.837	0.42	0.884	0.859	0.837	0.916	0.844	0.547
Likelihood	-72.41	-54.3		-58.96	-72.41	-82.99	-71.95	-22.67
F-statistic	776.2	10.14			776.2	6437	785.1	
Prob>F	0.000	0.010			0.000	0.000	0.000	

Robust standard errors in parentheses

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

[a] See variable definitions in Table SM.2.

Table SM.4a. Sensitivity of HS Results to Removing Control Variables, Below $\leq 4°C$

Specification	Preferred	Drop market control variable	Drop catastrophic control variable	Drop productivity control variable	Drop cross-sectional control variable	Noncatastrophic damages as dependent variable and drop catastrophic control variable
	(0)	(1)	(2)	(3)	(4)	(5)
VARIABLES[a]	D_new	D_new	D_new	D_new	D_new	Damage
t2	0.595**	0.417*	0.731***	0.595**	0.595**	0.482**
	(0.190)	(0.196)	(0.092)	(0.185)	(0.185)	(0.155)
mkt_t2	-0.622**		-0.757***	-0.535**	-0.433	-0.508**
	(0.226)		(0.150)	(0.193)	(0.248)	(0.195)
cat_t2	0.260	0.438		0.260	0.260	
	(0.267)	(0.254)		(0.259)	(0.259)	
prod_t2	0.113	-0.330	0.113		-0.075	0.113
	(0.125)	(0.200)	(0.122)		(0.194)	(0.123)
Cross	1.700***	0.924	1.700***	1.549***		1.700***
	(0.343)	(0.690)	(0.332)	(0.405)		(0.336)
Observations	21	21	21	21	21	19
R2	0.722	0.658	0.702	0.72	0.679	0.55
Adjusted R2	0.635	0.578	0.632	0.654	0.604	0.43
Likelihood	-44.36	-46.54	-45.1	-44.45	-45.86	-38.34
F-statistic	21.95	12.69	31.64	17.45	15.13	18.09
Prob>F	0.000	0.002	0.000	0.001	0.001	0.000

Robust standard errors in parentheses

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

[a] See variable definitions in Table SM.2.

Table SM.4b. Sensitivity of HS Results to Removing Control Variables, All Data

Specification	Preferred	Drop market control variable	Drop catastrophic control variable	Drop productivity control variable	Drop cross-sectional control variable	Non-catastrophic damages as dependent variable and drop catastrophic control variable
	(0)	(1)	(2)	(3)	(4)	(5)[a]
VARIABLES[a]	D_new	D_new	D_new	D_new	D_new	Damage
t2	0.318**	0.279***	0.642***	0.318***	0.318***	0.312***
	(0.102)	(0.081)	(0.049)	(0.100)	(0.100)	(0.085)
mkt_t2	-0.345*		-0.668***	0.020	-0.156	-0.339**
	(0.156)		(0.125)	(0.210)	(0.189)	(0.144)
cat_t2	0.362***	0.402***		0.362***	0.362***	
	(0.103)	(0.082)		(0.101)	(0.101)	
prod_t2	0.398	0.093	0.398		0.209	0.398
	(0.237)	(0.217)	(0.232)		(0.258)	(0.235)
Cross	1.700***	1.167*	1.700***	1.062		1.700***
	(0.331)	(0.568)	(0.323)	(0.688)		(0.327)
Observations	26	26	26	26	26	22
R2	0.869	0.866	0.844	0.865	0.864	0.487
Adjusted R2	0.837	0.841	0.816	0.84	0.839	0.374
Likelihood	-72.41	-72.69	-74.61	-72.79	-72.89	-55.64
F-statistic	776.2	1008	51.67	1010	1009	11.65
Prob>F	0	0	0	0	0	0

Robust standard errors in parentheses

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

a See variable definitions in Table SM.2.

Table SM.5a. NM Regressions without Duplicates, by Temperature Limit

Specification	Damages for temp. increases≤4◦C		Damages for temp. increases≤8◦C		Damages for all temp. increases	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES[b]	D_new	D_new	D_new	D_new	D_new	D_new
t2		0.678**		0.270**		0.151*
		(0.220)		(0.105)		(0.072)
mkt_t2		-0.604*		-0.166		-0.047
		(0.316)		(0.214)		(0.209)
cat_t2		-0.390		0.018		0.137
		(0.497)		(0.475)		(0.492)
cs_t2		0.331		0.739		0.858
		(0.583)		(0.571)		(0.594)
T2_new	0.415**		0.252**		0.157**	
	(0.164)		(0.091)		(0.067)	
Observations	13	13	15	15	16	16
R2	0.347	0.6	0.355	0.49	0.267	0.397
Adjusted R2	0.293	0.422	0.309	0.305	0.218	0.195
Likelihood	-36.05	-32.87	-41.76	-40	-45.62	-44.07
F-statistic	6.389	3.373	7.706	2.643	5.466	1.972
Prob>F	0.0265	0.06	0.0149	0.091	0.0337	0.163

Standard errors in parentheses

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

[a] The results in Specification (5) do not match Specification (7) in Table 3, as we do not reintroduce data errors in the data (Nordhaus 1994b; Kemfert 2001, 2002; Tol 2013; Hambel et al. 2015) as we did in the replication experiment. The number of observations drops from 18 to 16 because we do not observe the necessary data in Hambel et al. (2015) and we drop one estimate because of duplication (the above result is robust to including the duplicate estimate).

[b] See variable definitions in Table SM.2.

Table SM 5b. NM Regressions without Duplicates, by Temperature Limit and Weighting Assumption

Specification	Drop NM Weights				Drop NM Weights and Add Inverse Temperature Weight			
	Damages for temp. increases≤4◦C		Damages for all temp. increases		Damages for temp. increases≤4◦C		Damages for all temp. increases	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES[a]	D_new	D_new	D_new	D_new	D_new	D_new	D_new	D_new
t2		0.682**		0.083**		0.723**		0.113
		(0.287)		(0.038)		(0.317)		(0.073)
mkt_t2		-0.597		0.055		-0.637		0.015
		(0.358)		(0.124)		(0.389)		(0.166)
cat_t2		-0.394		0.205		-0.435		0.175
		(0.648)		(0.649)		(0.699)		(0.662)
cs_t2		0.327		0.926***		0.237		0.847**
		(0.363)		(0.251)		(0.411)		(0.285)
T2_new	0.533**		0.102*		0.474**		0.151*	
	(0.200)		(0.049)		(0.201)		(0.080)	
Observations	13	13	16	16	13	13	16	16
R2	0.371	0.747	0.228	0.656	0.317	0.68	0.192	0.563
Adjusted R2	0.319	0.635	0.177	0.542	0.26	0.538	0.138	0.417
Likelihood	-38.75	-32.83	-49.24	-42.77	-35.31	-30.38	-45.07	-40.15
F-statistic	7.084	6.648	4.437	5.732	5.567	4.777	3.557	3.864
Prob>F	0.0207	0.00897	0.0524	0.00812	0.0361	0.0241	0.0788	0.0305

Standard errors in parentheses

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

[a] See variable definitions in Table SM.2.

Table SM 5c. NM Regressions without Duplicates and Excluding Catastrophic Impacts, by Temperature Limit

Specification	Damages for temp. increases≤4◦C		Damages for temp. increases≤8◦C		Damages for all temp. increases	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES[a]	damage	damage	damage	damage	damage	damage
t2		0.564**		0.261**		0.150**
		(0.201)		(0.098)		(0.069)
mkt_t2		-0.490		-0.158		-0.046
		(0.306)		(0.205)		(0.201)
cs_t2		0.445		0.748		0.860
		(0.586)		(0.549)		(0.571)
T2_new	0.393**		0.245**		0.153**	
	(0.166)		(0.091)		(0.067)	
Observations	13	13	15	15	16	16
R2	0.318	0.531	0.34	0.476	0.259	0.386
Adjusted R2	0.261	0.391	0.293	0.344	0.209	0.245
Likelihood	-36.23	-33.79	-41.82	-40.09	-45.59	-44.07
F-statistic	5.595	3.778	7.207	3.626	5.23	2.728
Prob>F	0.036	0.048	0.018	0.045	0.037	0.087

Standard errors in parentheses

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

[a] See variable definitions in Table SM.2.

Table SM 6. Sensitivity of NM and HS to Outliers Using MM-Regression, by Temperature Level

Specification	Nordhaus and Moffat (2017)				Howard and Sterner (2017)	
	Damages for temp. increases ≤ 4°C		Damages for all temp. increases		Damages for temp. increases ≤ 4°C	Damages for all temp. increases
	(1)	(2)[a]	(3)	(4)[a]	(5)	(6)
VARIABLES[b]	D_new	damage	D_new	damage	D_new	D_new
t2		0.188		0.179***	0.404*	0.231***
		(0.205)		(0.0171)	(0.209)	(0.00926)
mkt_t2		-0.109		-0.0924***	-0.423	-0.250*
		(0.206)		(0.0243)	(0.233)	(0.114)
cs_t2		0.0731		0.0821***		
		(0.205)		(0.0189)		
T2_new	0.200**		0.185***			
	(0.0701)		(0.0187)			
cat_t2					0.359	0.376***
					(0.596)	(0.00679)
prod_t2					0.106	0.146
					(0.121)	(0.119)
Cross					1.604**	1.605***
					(0.493)	(0.474)
Observations	14	14	17	17	21	26

Robust standard errors in parentheses

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

[a] Rerun with noncatastrophic damages because negative catastrophic impacts lead to high noncatastrophic impact estimates.

[b] See variable definitions in Table SM.2.

Table SM 7. Rerunning the HS Model after Redefining Catastrophic Impacts

Specification	Damages for temp. increases ≤ 4°C			Damages for all temp. increases		
	HS Catastrophic Definition	Risk Premium Definition		HS Catastrophic Definition	Risk Premium Definition	
	Weighted Least Squares	Weighted Least Squares	MM-regression	Weighted Least Squares	Weighted Least Squares	MM-regression
VARIABLES[a]	D_new	D_new	D_new	D_new	D_new	D_new
t2	0.595**	0.613***	0.385*	0.318**	0.578***	0.229***
	(0.190)	(0.178)	(0.174)	(0.102)	(0.053)	(0.00962)
mkt_t2	-0.622**	-0.639**	-0.403*	-0.345*	-0.605***	-0.247*
	(0.226)	(0.216)	(0.204)	(0.156)	(0.129)	(0.112)
cat_t2	0.260	0.109	0.177	0.362***	0.143**	0.329**
	(0.267)	(0.174)	(0.191)	(0.103)	(0.061)	(0.127)
prod_t2	0.113	0.113	0.105	0.398	0.398	0.145
	(0.125)	(0.125)	(0.120)	(0.237)	(0.237)	(0.117)
Cross	1.700***	1.700***	1.586**	1.700***	1.700***	1.582**
	(0.343)	(0.343)	(0.521)	(0.331)	(0.331)	(0.507)
Observations	21	21	21	26	26	26
R2	0.72	0.68		0.87	0.86	
Adjusted R2	0.64	0.58		0.84	0.83	
Likelihood	-44.36	-43.47		-72.41	-69.37	
F-statistic	21.95	490.3		776.2	804	
Prob>F	0.000	0.000		0.000	0.000	

Robust standard errors in parentheses

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

[a] See variable definitions in Table SM.2.

Table SM 8a. Regressions Using Combined NM and HS Data, by Methodological Assumption

Specification	Selection Assumptions						Data Entry Assumption	
	Preferred	Include NM Duplicates	Stronger Global Definition: Drop Haneman n (2018)	Weaker Implementation of NM Selection Criteria: Include Cline (1992)	Only WTP Estimates	NM Data Alone	Redefine Catastrophic More Generally Following HS	Redefine Quality
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES[a]	D_new	D_new	D_new	D_new	D_new	damage	D_new	D_new
t2	0.594***	0.593***	0.594***	0.308***	0.594***	0.191***	0.334**	0.629***
	(0.033)	(0.033)	(0.033)	(0.073)	(0.032)	(0.028)	(0.132)	(0.027)
mkt_t2	-0.536***	-0.545***	-0.536***	-0.252**	-0.536***	-0.193***	-0.278*	-0.550***
	(0.073)	(0.066)	(0.074)	(0.098)	(0.073)	(0.040)	(0.147)	(0.154)
cat_t2	0.180	0.027	0.218	0.312***	0.180		0.286**	0.126
	(0.234)	(0.252)	(0.303)	(0.074)	(0.230)		(0.132)	(0.178)
prod_t2	0.187	0.129	0.187	0.189	0.187	0.120***	0.189	0.298
	(0.148)	(0.107)	(0.148)	(0.145)	(0.146)	(0.027)	(0.146)	(0.229)
cs_t2	0.366***	0.325***	0.366***	0.630***		0.769***	0.626***	0.331***
	(0.033)	(0.033)	(0.033)	(0.073)		(0.028)	(0.132)	(0.027)
Observations	29	30	28	32	27	14	29	29
R2	0.93	0.921	0.93	0.907	0.932	0.705	0.943	0.855
Adjusted R-squared	0.916	0.905	0.915	0.89	0.92	0.587	0.931	0.825
Likelihood	-74.55	-74.21	-72.86	-86.49	-69.15	-23.87	-71.67	-79.36
Damages without productivity								
Noncatastrophic damages at 3 degrees	6.7%	6.7%	6.7%	3.5%	6.7%	2.1%	3.8%	7.1%
Noncatastrophic damages at 6 degrees	26.7%	26.7%	26.7%	13.9%	26.7%	8.6%	15.0%	28.3%
Total damages at 3 degrees	8.3%	6.9%	8.6%	6.3%	8.3%	3.9%	6.3%	8.2%
Total damages at 6 degrees	33.2%	27.7%	34.6%	25.1%	33.2%	15.5%	25.3%	32.8%
Damages with productivity								
Noncatastrophic damages at 3 degrees	8.8%	8.1%	8.8%	5.6%	8.8%	3.5%	5.9%	10.4%
Noncatastrophic damages at 6 degrees	35.1%	32.5%	35.1%	22.4%	35.1%	14.0%	23.5%	41.7%

Total damages at 3 degrees	10.4%	8.4%	10.7%	8.4%	10.4%	5.2%	8.5%	11.6%
Total damages at 6 degrees	41.6%	33.5%	43.0%	33.6%	41.6%	20.9%	33.8%	46.3%

Robust standard errors in parentheses

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

a See variable definitions in Table SM.2.

Table SM 8b. Regressions Using Combined NM and HS Data, by Methodological Assumption

Specification	Preferred	Control variables				Weights	Outliers	
		No control	Original HS	Drop CS	Extended HS[a]	No Hetero-skedastic Weight	Drop above 4 degrees	MM-regression
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES[b]	D_new	D_new	D_new	D_new	D_new	D_new	D_new	MM
t2	0.594***	0.583***	0.594***	0.595***	0.594***	0.600***	0.613***	0.249**
	(0.033)	(0.038)	(0.032)	(0.031)	(0.033)	(0.017)	(0.199)	(0.098)
mkt_t2	-0.536***		-0.384***	-0.538***	-0.562***	-0.570***	-0.556**	-0.208*
	(0.073)		(0.121)	(0.072)	(0.053)	(0.029)	(0.210)	(0.105)
cat_t2	0.180		0.180	0.179	0.180	0.203	0.161	0.384
	(0.234)		(0.229)	(0.228)	(0.239)	(0.232)	(0.232)	(0.765)
cs_t2	0.366***		0.366***		0.114	0.409***	0.347	0.018
	(0.033)		(0.032)		(0.384)	(0.017)	(0.199)	(0.098)
prod_t2	0.187			0.187	0.212	0.257	0.036	0.067
	(0.148)			(0.145)	(0.140)	(0.165)	(0.068)	(0.052)
cross_t2					0.251			
					(0.382)			
Observations	29	29	29	29	29	29	24	29
R2	0.93	0.91	0.93	0.929	0.93	0.973	0.72	0.535
Adjusted R-squared	0.916	0.906	0.918	0.918	0.912	0.967	0.646	
Likelihood	-74.55	-78.31	-74.68	-74.82	-74.52	-84.67	-52.02	
Damages without productivity								
Noncatastrophic damages at 3 degrees	6.7%	6.6%	6.7%	6.7%	6.7%	6.8%	6.9%	2.8%
Noncatastrophic damages at 6 degrees	26.7%	26.2%	26.7%	26.8%	26.7%	27.0%	27.6%	11.2%
Total damages at 3 degrees	8.3%	-	8.3%	8.3%	8.3%	8.6%	8.3%	6.3%
Total damages at 6 degrees	33.2%	-	33.2%	33.2%	33.2%	34.3%	33.4%	25.0%
Damages with productivity								
Noncatastrophic damages at 3 degrees	8.8%	-	-	8.8%	9.1%	9.6%	7.3%	3.6%
Noncatastrophic damages at 6 degrees	35.1%	-	-	35.2%	36.3%	38.6%	29.2%	14.2%
Total damages at 3 degrees	10.4%	-	-	10.4%	10.7%	11.5%	8.8%	7.0%

Total damages at 6 degrees	41.6%	-	-	41.6%	42.8%	45.9%	35.0%	28.0%

Robust standard errors in parentheses

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

[a] We slightly modify the HS extended specification to interact the cross-sectional indicator variable with temperature-squared.

[b] See variable definitions in Table SM.2.

Table SM 9. Standard Deviation Regressions, by Functional Form[a]

Specification	Quadratic				Linear				Square-Root			
	All SD Est.		Observed SD Est.		All SD Est.		Observed SD Est.		All SD Est.		Observed SD Est.	
	All Temp	≤°4	All Temp	≤°4	All Temp	≤°4	All Temp	≤°4	All Temp	≤°4	All Temp	≤°4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	SD	SD	SD	SD	SD	SD	SD	SD	SD	SD	SD	SD
t2	0.564	0.651	1.187***	0.837**								
	(0.317)	(0.318)	(0.097)	(0.194)								
t					2.625	1.730	4.264*	2.368*				
					(1.179)	(0.986)	(1.288)	(0.629)				
t_sqrt									4.680	2.610	6.580*	3.548
									(1.925)	(1.673)	(1.917)	(1.383)
Observations	10	8	7	6	10	8	7	6	10	8	7	6
R2	0.544	0.628	0.928	0.787	0.526	0.595	0.728	0.79	0.475	0.549	0.592	0.723
Adjusted R-squared	0.493	0.575	0.916	0.744	0.474	0.537	0.683	0.748	0.417	0.485	0.524	0.668
Liklihood	-34.49	-20.56	-18.78	-14.6	-34.68	-20.89	-23.42	-14.55	-35.19	-21.32	-24.84	-15.38
AIC	71.0	43.1	39.6	31.2	71.4	43.8	48.8	31.1	72.4	44.6	51.7	32.8
Prob min est info loss	43%	99%	46%	40%	36%	1%	33%	42%	21%	0%	21%	18%

Robust standard errors in parentheses

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

[a] See variable definitions in Table SM.2. Additionally, t is a linear temperature term, while t_sqrt is the square root of temperature.

Table SM 10. Quality Effects and Extended Fixed Effect Regressions Using Combined NM and HS Data, by Methodological Assumption

Data[a]	**Total Damages**				**Noncatastrophic Damages**[b]			
Estimator	Quality effects		Fixed effects		Quality effects		Fixed effects	
Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES[c]	D_new	D_new	D_new	D_new	damage	damage	Damage	damage
t2	0.611***	0.611***	0.703***	0.984***	0.611***	0.611***	0.638***	0.989***
	(0.006)	(0.007)	(0.094)	(0.132)	(0.006)	(0.006)	(0.098)	(0.119)
mkt_t2	-0.063	-0.515***	-0.267	-0.830	-0.063	-0.514***	-0.202	-0.628
	(0.271)	(0.106)	(0.192)	(0.604)	(0.266)	(0.104)	(0.191)	(0.715)
cat_t2	0.412***	0.412**	-0.003	-0.072				
	(0.146)	(0.149)	(0.133)	(0.094)				
prod_t2		0.474		0.083		0.474		0.100
		(0.300)		(0.577)		(0.295)		(0.597)
qual_t2				0.824				0.482
				(0.468)				(0.537)
age_t2				-0.685**				-0.793**
				(0.299)				(0.281)
Observations	27	27	27	27	25	25	25	25
R2	0.989	0.989	0.568	0.641	0.99	0.99	0.526	0.585
Adjusted R2	0.988	0.988	0.514	0.539	0.989	0.989	0.485	0.481
Likelihood	-78.74	-78.56	-58.3	-55.78	-75.19	-75.02	-53.73	-52.09

Robust standard errors in parentheses

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

[a] Section 5 only analyzes willingness to pay studies dropping compensating estimates

[b] When possible, all catastrophic components are removed from estimates. Nordhaus (2014) is the sole study for which catastrophic component of damages cannot be removed, such that its corresponding estimates are dropped.

[c] See variable definitions in Table SM.2.

Table SM 11. Quality, Random, and Fixed Effect Models, by Methodological Assumption

Weights with Study	**Preferred - Weights Sum to 1 within Studies**	**All Estimates Given Equal Weight**				
Estimator	Quality Effects		Random Effects		Fixed Effects	
Specification	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES[a]	D_new	D_new	D_new	D_new	D_new	D_new
t2	0.611***	0.643***	0.668***	0.984***	0.680***	0.974***
	(0.007)	(0.051)	(0.144)	(0.288)	(0.143)	(0.287)
mkt_t2	-0.515***	-0.542***	-0.442	-0.614	-0.542	-0.784
	(0.106)	(0.120)	(0.529)	(0.599)	(0.496)	(0.557)
cat_t2	0.412**	0.189	0.037	-0.054	0.025	-0.057
	(0.149)	(0.231)	(0.496)	(0.502)	(0.492)	(0.498)
prod_t2	0.474	0.435	0.146	-0.069	0.269	0.051
	(0.300)	(0.295)	(0.560)	(0.580)	(0.525)	(0.546)
qual_t2				0.483		0.612
				(0.557)		(0.542)
age_t2				-0.726		-0.687
				(0.559)		(0.555)
Observations	27	27	27	27	27	27
R2	0.989	0.957				
Adjusted R2	0.988	0.95				
Likelihood	-78.56	-94.71				
tau2			0.867	0.825		
Cochran's Q			15.44	13.25		
chi2			0.884	0.746		
I-squared			0	0		

Robust standard errors in parentheses

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

[a] See variable definitions in Table SM.2.

Table SM 12. Quality Effects Regressions Using Combined NM and HS Data, by Quality Weight Assumption

Specification	Preferred	Drop estimate-specific quality weight (w_quality)	Assign 25% quality weight (w_quality =0.25) for poor quality studies	Assume weather-based studies are poor quality (w_quality =0.1)	Assume weather-based studies are good quality (w_quality =1.0	Assume science-based studies are poor quality (w_quality =0.1)
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES[a]	D_new	D_new	D_new	D_new	D_new	D_new
t2	0.611***	0.611***	0.611***	0.611***	0.611***	0.613***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.012)
mkt_t2	-0.515***	-0.473***	-0.515***	-0.516***	-0.514***	-0.503***
	(0.106)	(0.156)	(0.106)	(0.105)	(0.107)	(0.123)
cat_t2	0.412**	0.412**	0.412**	0.412**	0.412**	0.111
	(0.149)	(0.149)	(0.149)	(0.149)	(0.149)	(0.209)
prod_t2	0.474	0.357	0.474	0.218	0.559*	0.459
	(0.300)	(0.307)	(0.300)	(0.239)	(0.284)	(0.306)
Observations	27	27	27	27	27	27
R2	0.989	0.986	0.989	0.991	0.988	0.953
Adjusted R2	0.988	0.983	0.988	0.99	0.986	0.945
Likelihood	-78.56	-79.15	-78.56	-76.63	-79.81	-74.65

Robust standard errors in parentheses

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

a See variable definitions in Table SM.2.

Table SM 13. Quality Effects Regressions Using Combined NM and HS Data, by Age Weight Assumption

Specification	Preferred	Drop age weight: g(Age)=0	Logistic age penalty function	Redefine age for Hanemann (2008) and Weitzman (2012) using papers instead of Ackerman and Stanton
	(1)	(2)	(3)	(4)
VARIABLES[a]	D_new	D_new	D_new	D_new
t2	0.611***	0.604***	0.611***	0.611***
	(0.007)	(0.010)	(0.006)	(0.007)
mkt_t2	-0.515***	-0.520***	-0.520***	-0.515***
	(0.106)	(0.090)	(0.099)	(0.105)
cat_t2	0.412**	0.397**	0.391**	0.418***
	(0.149)	(0.166)	(0.165)	(0.146)
prod_t2	0.474	0.444	0.437	0.475
	(0.300)	(0.297)	(0.301)	(0.299)
Observations	27	27	27	27
R2	0.989	0.981	0.989	0.989
Adjusted R2	0.988	0.978	0.987	0.987
Likelihood	-78.56	-82.46	-78.98	-78.24

Robust standard errors in parentheses

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

[a] See variable definitions in Table SM.2.

Table SM 14. Quality Effects Regressions Using Combined NM and HS Data, by Duplication Weight Assumption

Specification	Preferred	100% Weight of Duplicates in NM	25% Weight of Duplicates in NM	10% Weight of Duplicates in NM
	(1)	(2)	(3)	(4)
VARIABLES[a]	D_new	D_new	D_new	D_new
t2	0.611***	0.611***	0.611***	0.611***
	(0.007)	(0.007)	(0.007)	(0.007)
mkt_t2	-0.515***	-0.511***	-0.514***	-0.514***
	(0.106)	(0.111)	(0.107)	(0.106)
cat_t2	0.412**	0.337	0.392**	0.404**
	(0.149)	(0.210)	(0.166)	(0.156)
prod_t2	0.474	0.348	0.436	0.458
	(0.300)	(0.309)	(0.306)	(0.303)
Observations	27	27	27	27
R2	0.989	0.988	0.989	0.989
Adjusted R2	0.988	0.986	0.987	0.987
Likelihood	-78.56	-77.9	-78.45	-78.52

Robust standard errors in parentheses

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

[a] See variable definitions in Table SM.2.

Table SM 15. Quality Effects Regressions Using Combined NM and HS Data, by Multiple Estimates per Study Weight and Multiple Studies per Methodology Weight Assumption

Preferred	Preferred	Move multiple estimates per study (1/n) and multiple studies per methodology (1/N) weights to 2nd step of QE	Drop multiple estimates per study weight (1/n)	Drop multiple studies per methodology weight (1/N)	Drop multiple estimates per study (1/n) and the multiple studies per methodology weights (1/N)	Replace multiple estimates per study weight (1/n) with relative within-methodology weight (1/ sum_Beta)	Split enumerative and CGE into two separate methods to redefine the multiple studies per methodology weight (1/N)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES[a]	D_new	D_new	D_new	D_new	D_new	D_new	D_new
t2	0.611***	0.611***	0.643***	0.611***	0.643***	0.611***	0.611***
	(0.007)	(0.007)	(0.051)	(0.007)	(0.051)	(0.007)	(0.007)
mkt_t2	-0.515***	-0.475***	-0.514***	-0.474***	-0.513***	-0.494***	-0.515***
	(0.106)	(0.156)	(0.153)	(0.156)	(0.154)	(0.132)	(0.110)
cat_t2	0.412**	0.410**	0.189	0.230	0.115	0.345*	0.376**
	(0.149)	(0.150)	(0.230)	(0.222)	(0.200)	(0.188)	(0.172)
prod_t2	0.474	0.428	0.402	0.442	0.417	0.522*	0.355
	(0.300)	(0.321)	(0.310)	(0.320)	(0.310)	(0.298)	(0.295)
Observations	27	27	27	27	27	27	27
R2	0.989	0.989	0.957	0.976	0.947	0.982	0.987
Adjusted R2	0.988	0.987	0.949	0.972	0.938	0.979	0.985
Likelihood	-78.56	-77.05	-93.52	-75	-86.45	-76.69	-77.93

Robust standard errors in parentheses

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

[a] See variable definitions in Table SM.2.

Table SM 16. Quality Effects Regressions Using Combined NM and HS Data, by Estimation Methodology Weight

Specification	Preferred	Alpha 1 - Nordhaus's implicit weight	Alpha_2 - Howard and Sterner's implicit weight	Alpha_3 - Enumerative Only	Alpha 4 - No scientific or catastrophic	Alpha 5 - No Survey	Alpha 6 - No Methodology Weight
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES[a]	D_new	D_new	D_new	D_new	damage	D_new	D_new
t2	0.611***	0.682**	0.611***	0.244*	0.608**	0.611***	0.611***
	(0.007)	(0.307)	(0.007)	(0.118)	(0.256)	(0.006)	(0.007)
mkt_t2	-0.515***	-0.601*	-0.485***	-3.421***	-0.501*	-0.517***	-0.486***
	(0.106)	(0.321)	(0.142)	(0.025)	(0.282)	(0.104)	(0.142)
cat_t2	0.412**	-0.179	0.196	0.259		0.413**	0.231
	(0.149)	(0.340)	(0.223)	(0.209)		(0.152)	(0.222)
prod_t2	0.474	0.235	0.499	3.285***	0.461	0.477	0.456
	(0.300)	(0.240)	(0.306)	(0.118)	(0.302)	(0.304)	(0.313)
Observations	27	23	27	10	23	22	27
R2	0.989	0.636	0.967	0.749	0.623	0.991	0.976
Adjusted R2	0.988	0.559	0.961	0.582	0.566	0.989	0.972
Likelihood	-78.56	-59.24	-75.38	-15.82	-62.9	-64.9	-75.57

Robust standard errors in parentheses

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

[a] See variable definitions in Table SM.2.

Table SM 17. Quality Effects Regressions Using Combined NM and HS Data Set with Step-by-Step Approximation of NM Weights, by Weighting Assumption

Specification	Preferred	Replace the quality weights with w_Q	Replace Nordhaus' quality weights without duplicates	Allow for duplicates	Down weight Science and Panel to 10%
	(1)	(2)	(3)	(4)	(5)
VARIABLES[a]	D_new	D_new	D_new	D_new	D_new
t2	0.611***	0.599***	0.601***	0.601***	0.555***
	(0.007)	(0.015)	(0.009)	(0.009)	(0.068)
mkt_t2	-0.515***	-0.471***	-0.557***	-0.567***	-0.513***
	(0.106)	(0.151)	(0.046)	(0.029)	(0.080)
cat_t2	0.412**	0.195	0.191	0.032	-0.022
	(0.149)	(0.234)	(0.242)	(0.262)	(0.165)
prod_t2	0.474	0.402	0.263	0.185	0.124
	(0.300)	(0.321)	(0.212)	(0.154)	(0.088)
Observations	27	27	27	27	27
R2	0.989	0.962	0.981	0.98	0.895
Adjusted R2	0.988	0.955	0.978	0.976	0.877
Likelihood	-78.56	-75.79	-70.37	-69.21	-66.77

Robust standard errors in parentheses

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

[a] See variable definitions in Table SM.2.

Table SM 18. Quality Effects Regressions Using Combined NM and HS Data, by Minimum Weight Criterion

Specification	Preferred	Minimum weight of 10%	Minimum weight of 25%
	(1)	(2)	(3)
VARIABLES[a]	D_new	D_new	D_new
t2	0.611***	0.609***	0.612***
	(0.007)	(0.006)	(0.014)
mkt_t2	-0.515***	-0.551***	-0.560***
	(0.106)	(0.071)	(0.089)
cat_t2	0.412**	0.414**	0.345*
	(0.149)	(0.149)	(0.189)
prod_t2	0.474	0.334	0.337
	(0.300)	(0.241)	(0.247)
Observations	27	27	27
R2	0.989	0.986	0.969
Adjusted R-squared	0.988	0.983	0.964
Likelihood	-78.56	-79.12	-82.68

Robust standard errors in parentheses

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

[a] See variable definitions in Table SM.2.

Table SM 19. Quality Effects Regressions Using Combined NM and HS Data, by Functional Form of Standard Deviation

Preferred	Preferred	Drop estimates above 4 degrees to address nonlinear standard errors	Standard errors are quadratic function in temperature	Standard errors are square-root function in temperature
	(1)	(2)	(3)	(4)
VARIABLES[a]	D_new	D_new	D_new	D_new
t2	0.611***	0.956***	0.611***	0.611***
	(0.007)	(0.170)	(0.007)	(0.007)
mkt_t2	-0.515***	-0.869***	-0.501***	-0.520***
	(0.106)	(0.195)	(0.123)	(0.098)
cat_t2	0.412**	0.109	0.412**	0.411**
	(0.149)	(0.209)	(0.149)	(0.149)
prod_t2	0.474	-0.021	0.463	0.476
	(0.300)	(0.098)	(0.306)	(0.298)
Observations	27	22	27	27
R2	0.989	0.93	0.989	0.989
Adjusted R2	0.988	0.915	0.988	0.987
Likelihood	-78.56	-46.49	-77.96	-78.55

Robust standard errors in parentheses

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

[a] See variable definitions in Table SM.2.

Table SM 20. Quality Effects Regressions Using Combined NM and HS Data, by Definition of Catastrophic Impact

Specification	Preferred (Narrower) Definition of Catastrophic	HS (Broader) Definition of Catastrophic
	(1)	(2)
VARIABLES[a]	D_new	D_new
t2	0.611***	0.677**
	(0.007)	(0.302)
mkt_t2	-0.515***	-0.580*
	(0.106)	(0.320)
cat_t2	0.412**	-0.064
	(0.149)	(0.302)
prod_t2	0.474	0.474
	(0.300)	(0.300)
Observations	27	27
R2	0.989	0.987
Adjusted R2	0.988	0.985
Likelihood	-78.56	-80.82

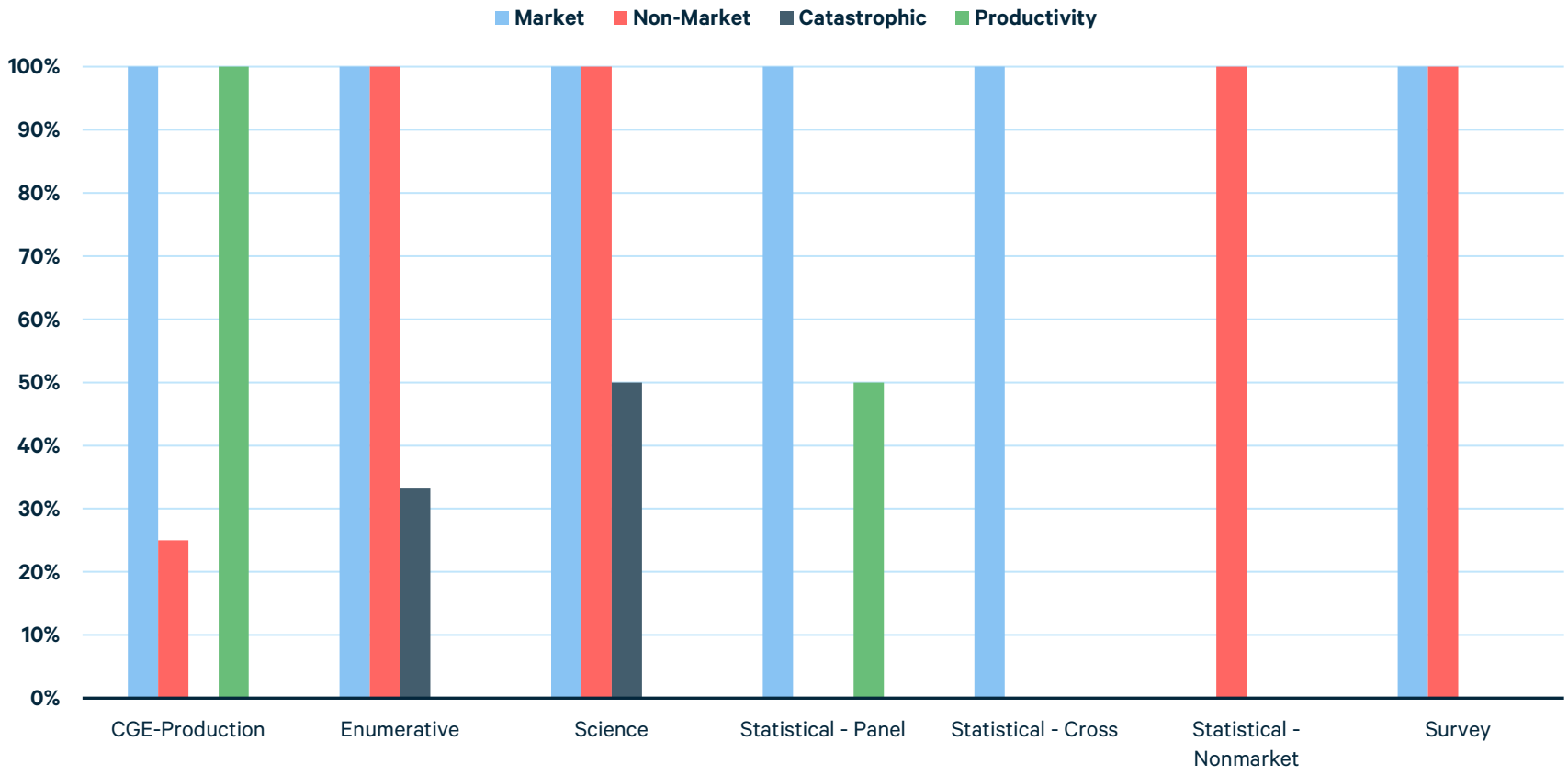
Robust standard errors in parentheses

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

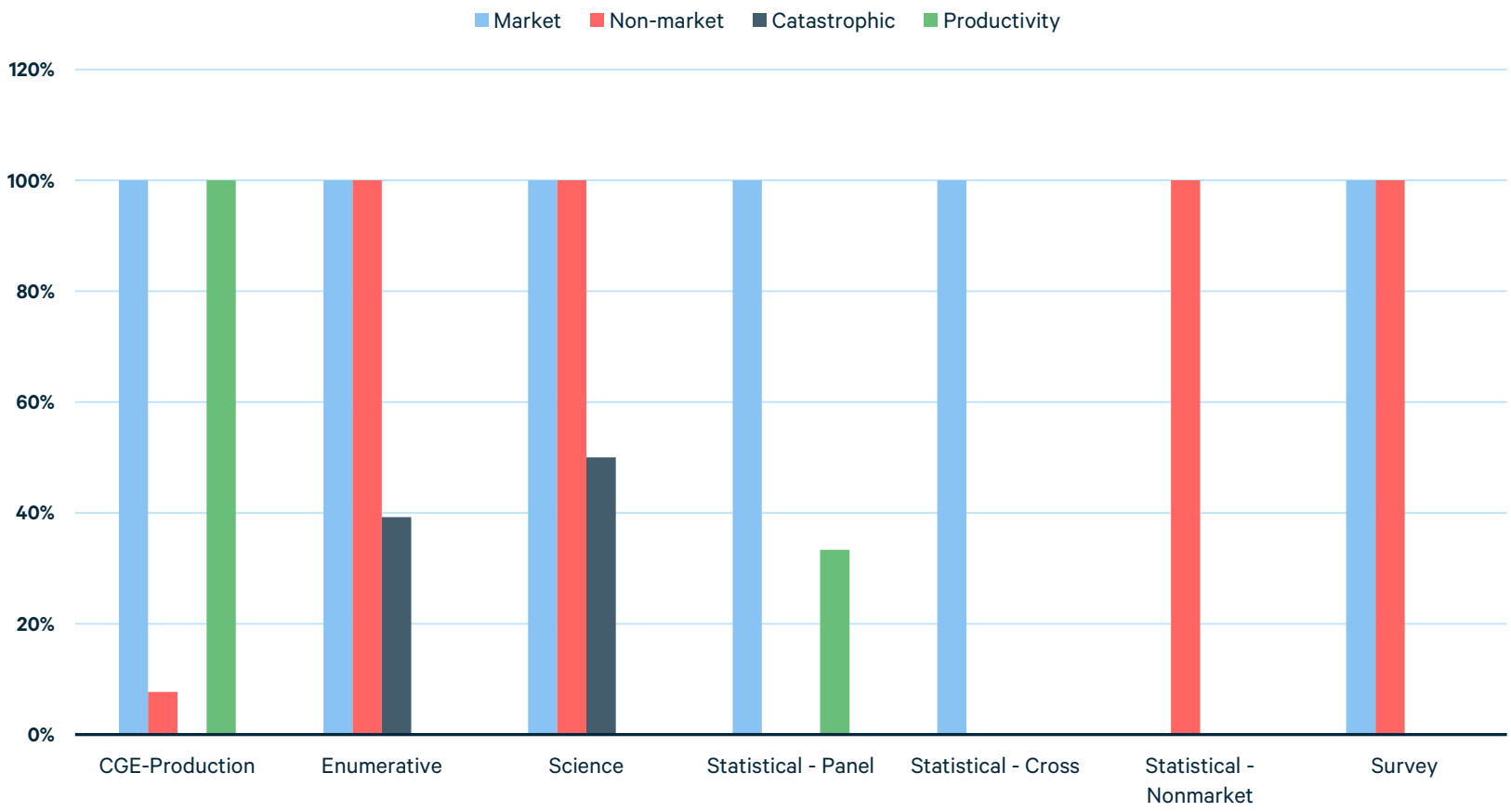
[a] See variable definitions in Table SM.2.

Figure SM 1. Share of Each Climate Impact Type Captured by Each Estimation Methodology for Climate Damages

A. Equal weighting



B. NM weighting



C. Preferred quality effects weight

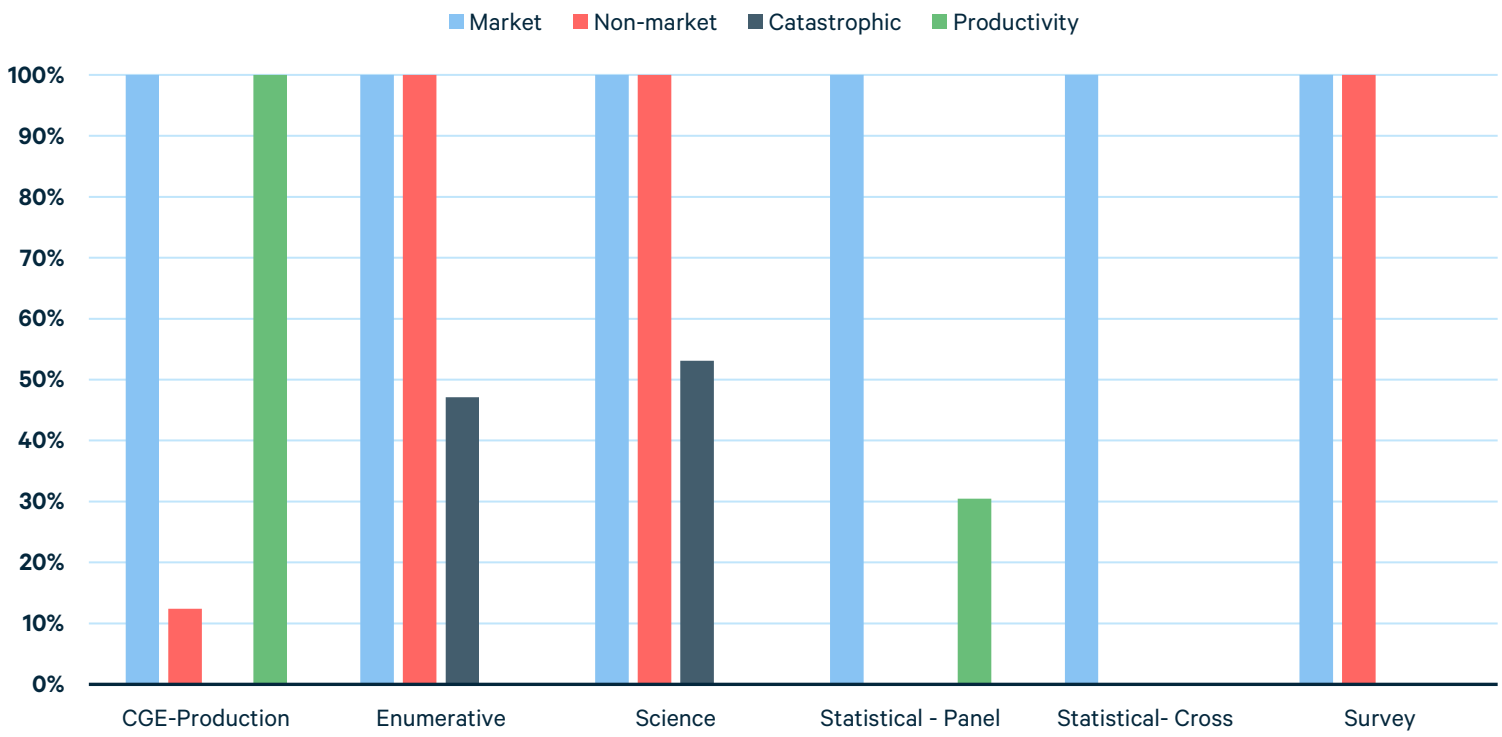


Figure SM 2. Effect of Dropping Each Estimate, by Control Variable

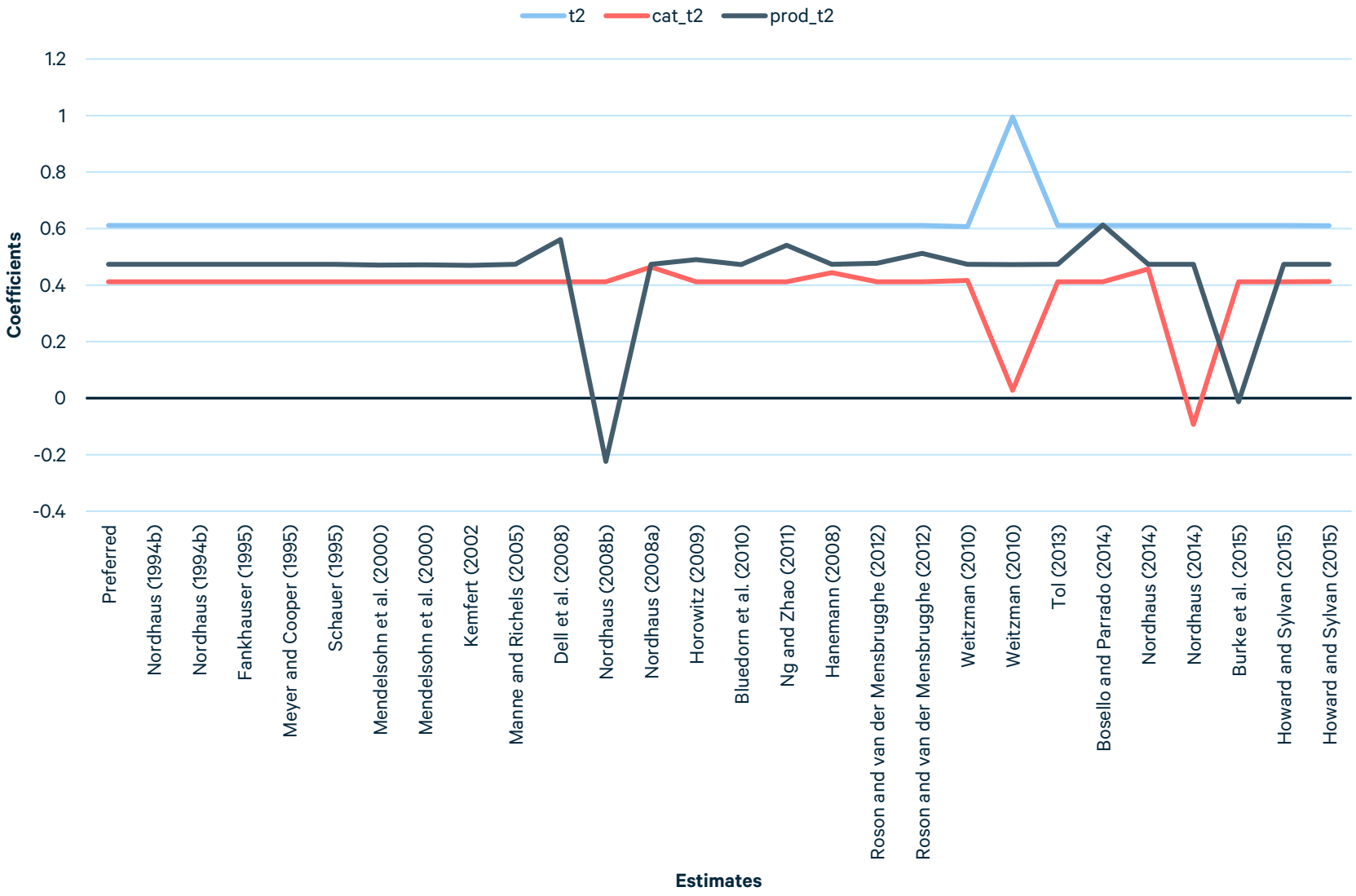


Figure SM.3.Effect of Dropping Each Study, by Control Variable

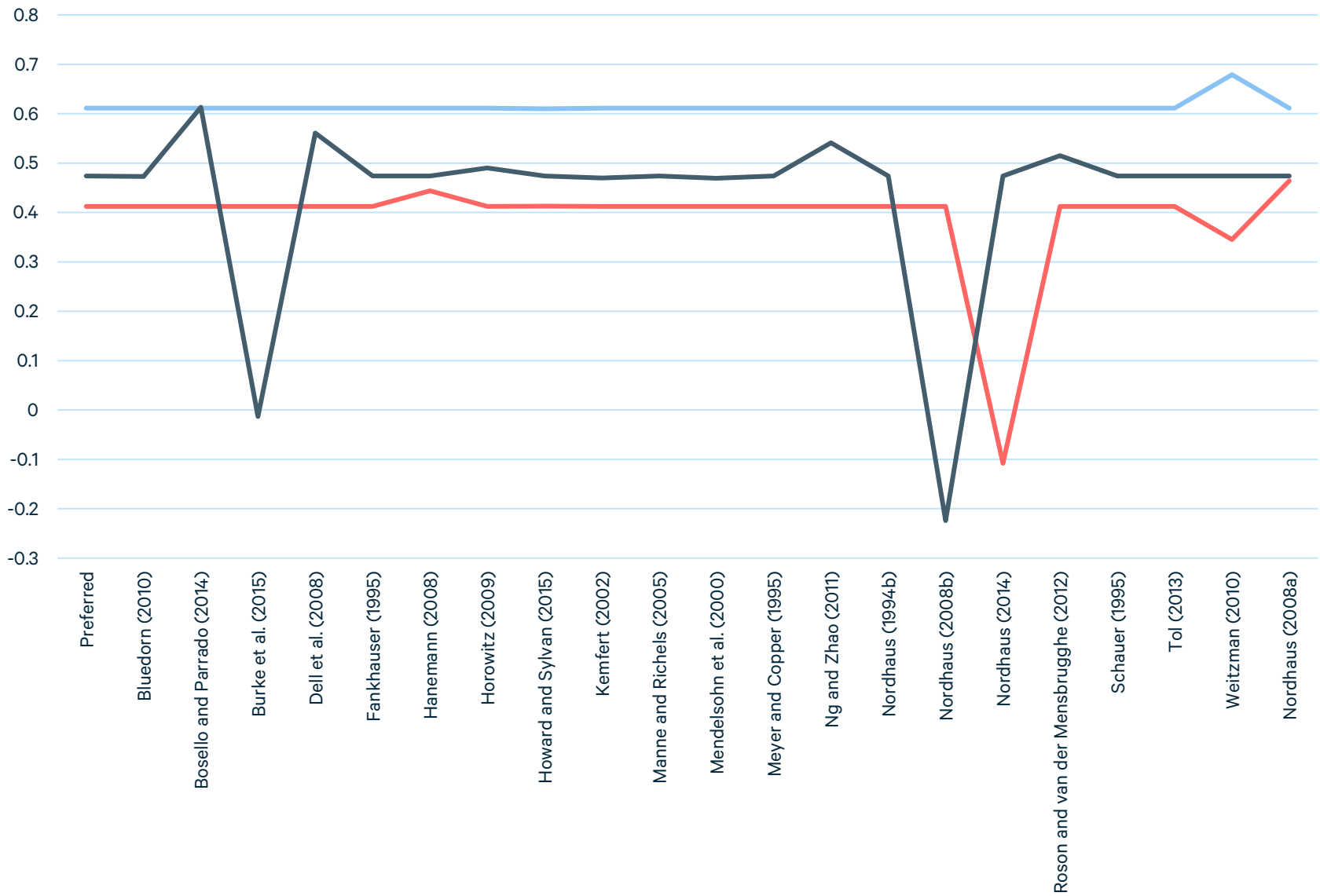
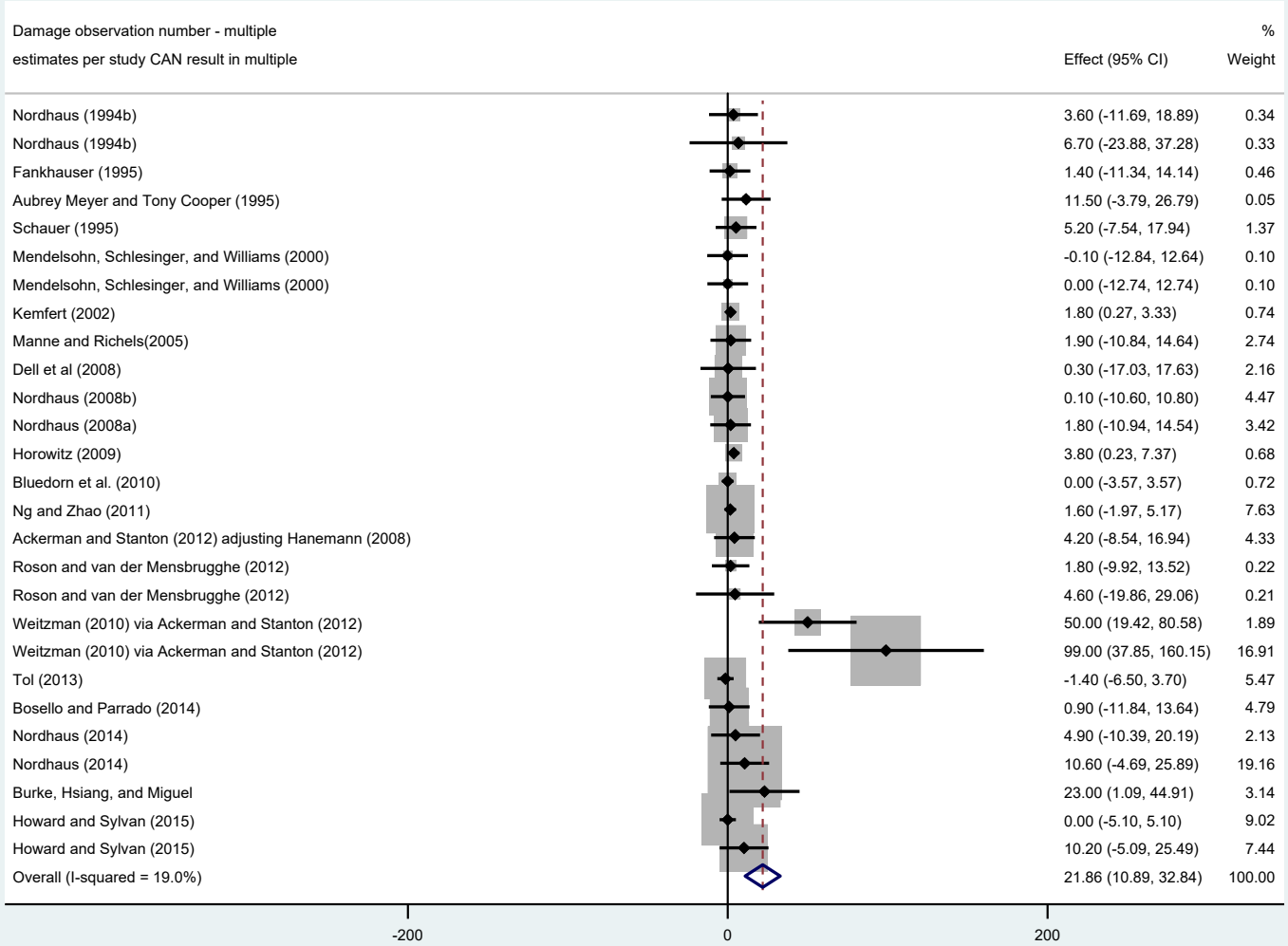


Figure SM 4. Forest Plot for First Stage of Preferred Specification of Quality Effects Estimator

A.) Equal weighting of studies within methodology



B. Proportional weighting of studies within methodology