AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Few and Not So Far Between: A Meta-analysis of Climate Damage Estimates

Peter H. Howard
Economics Director
Institute for Policy Integrity, New York University School of Law
HowardP@mercury.law.nyu.edu

and

Thomas Sterner
Professor of Environmental Economics
Department of Economics at the University of Gothenburg
thomas.sterner@economics.gu.se

Selected Paper prepared for presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, Massachusetts, July 31-August 2

Copyright 2016 by [authors]. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Abstract

In the United States, the social cost of carbon (SCC) is one of the foremost tools for calibrating the socially optimal approach for climate change policy. The SCC is estimated using climate-economic models with implicit temperature-damage relationships. Given the vast uncertainty surrounding climate impacts, meta-analyses of global climate damage estimates are a key tool for determining the relationship between temperature and climate damages, so as to communicate the current state of knowledge to model developers. Using a larger dataset than previously assembled in the literature, this paper highlights several methodological improvements that address bias present in previous meta-analyses of the temperature-damage relationship. Specifically, due to limited data availability, previous meta-analyses of global climate damages potentially suffered from multiple sources of bias: duplication bias, measurement error, omitted variable bias, and publication bias. By expanding our dataset (to include additional published and grey literature estimates), including methodological variables, and correcting the specification of temperature (to account for different reference periods), we are able to address and test for these biases. Estimating the relationship between temperature and climate damages using weighted least squares with cluster robust standard errors at the model level, we find strong evidence of duplicate bias. Using these results as an input in the DICE model – to update the DICE damage function – we determine that duplication and omitted variable bias significantly impact the damage-temperature relationship in past meta-analyses and the resulting SCC estimates. Focusing exclusively on non-catastrophic climate impacts, we find that the temperature-damage relationship estimated in Nordhaus and Sztorc (2013) is biased downwards by approximately 179% to 264%, depending on how climate change's impacts on productivity are treated. This implies a downward bias in DICE's SCC estimate by 203% to 314%, depending on the treatment of productivity. If we also consider catastrophic impacts, the potential bias in the SCC increases to 344% to 469%.

Few and Not So Far Between: A Meta-analysis of Climate Damage Estimates

Peter H Howard and Thomas Sterner

Climate change is one of the preeminent policy issues of our day, and the Paris agreement signals the end of denial. The agreement is still not clear on methods of implementation but there is a unique international understanding that action is now urgently needed. The social cost of carbon (SCC) is one of the primary tools for calibrating the socially optimal policy response, particularly in the U.S. The SCC is estimated using Integrated Assessment Models (IAMs), which capture the various steps in the climate and economic processes that translate a marginal unit of CO_2 emissions into an economic damage. While accuracy within each of these steps is necessary to precisely estimate the SCC, accurately calibrating the climate damage function – which translates a temperature change into a percentage change in GDP – is critical. Given the considerable uncertainty in climate impacts, meta-analyses of climate damage estimates are a key tool for depicting the relationship between temperature and climate damages, so as to communicate the current state of knowledge to model developers (Bergh and Button, 1999). By clarifying this relationship and the uncertainty underlying it, meta-analyses also explain to policymakers – in addition to climate-economic modelers – the best assessment of the risks that climate change poses to the global economic system and to human well-being.

Due to a dearth of global damage estimates, past meta-analyses (Newbold and Marten, 2014; Nordhaus and Sztorc, 2013; Tol, 2009; Tol, 2013; Tol, 2014) conducted relatively limited analyses of the temperature-damage relationship. It is fair to say that the first damage studies made the very reasonable choice of using simple "back of the envelope" estimates for damages. Gradually, the number of studies increased as methodologies improved, and authors were able to gather small datasets varying from 13 to 43 damage estimates (acquired from 13 to 17 early studies) as a basis for assigning damage values. However, we believe that these earlier analyses are quite limited and potentially even flawed since these datasets were characterized by non-independent observations. Many observations were estimated using similar methods and models, and very often undertaken by the same authors and thus verging on duplication.

The idea of using meta-analyses to summarize the state of knowledge is well established. However, a case in which a very limited number of authors are taking averages of their own earlier values can be problematic. In fact, 5 of 13 studies used in the regression by Nordhaus and Sztorc (2013) were Nordhaus' own earlier estimates, and an additional four were from the other two main modelers in the IAM literature. Given the small sample sizes available in the literature, authors of these prior meta-analyses simply regressed global climate damages (measured as a % of global GDP) on global temperature change (assumed to be relative to the pre-industrial period) and calculated basic OLS standard errors. In doing so, they did not account for duplication or dependency of damage estimates. The consequence is an exaggerated sense of knowledge about the temperature-damage relationship and a greater weighting of estimates from established authors in the field (who tend towards the enumerative strategy) away from new authors such as Burke et al (2015) (who tend toward statistical-based strategies); see also Sterner (2015).

These earlier meta-analyses find that global climate damages increase at an increasing rate, though they differ in the exact rate of change; see Figure 1. Specifically, the models predict a wide range of potential damage estimates from 1.9% to 17.3% of GDP for a 3 °C increase in global average surface temperatures. While these earlier estimates represent strong first steps in understanding the temperature-damage relationship, they are potentially biased due to these data limitations and the limiting effect of the estimation techniques employed. This is particularly true when they are used to predict the damages from higher temperatures, since there is a large sensitivity to the choice of functional form.

There are several sources of potential bias in previous empirical estimates of the temperature-damage relationship. First, because the data are dependent on one another due to similar underlying estimation methods, authors, and models, OLS is inefficient and the resulting standard errors are biased. More importantly, as indicated above, due to the common practices of citing earlier estimates and updating previous estimates in the literature, the resulting estimates suffer from multiple (duplicate) publication bias; this is the first paper to apply this meta-analysis concept fully in climate economics. Second, earlier estimates also suffer from omitted variable bias due to a failure to account for whether the underlying damage estimates captured non-market and catastrophic climate impacts in addition to market impacts and to control for whether market impact estimates included potential impacts on economic productivity. Third, earlier studies potentially suffer from publication bias by relying exclusively on published studies. Fourth, there are several citation errors with respect to damage estimates – in particular with respect to the measurement of temperature. Last, outlier damage estimates (corresponding to temperature changes above 3.2 °C) may have an undue impact on the estimated temperature-damage relationship over the critical time period of the 21st century. By dealing systematically with these potential biases, this paper identifies a significant downward bias in the literature of the effect of temperature change on global climate damages.

Our aim is thus to improve understanding of the temperature-damage relationship by advancing the meta-analysis techniques pioneered by Tol and Nordhaus. A meta-analysis is not necessarily synonymous with "truth", but clearly a meta-analysis should, if worth doing, be done well. By conducting an exhaustive search of the published and grey literature to assemble a larger dataset, we are able to conduct a meta-analysis using more data-intensive techniques that can more accurately capture this relationship and the uncertainty underlying it. Specifically, using weighted least squares and calculating cluster robust standard errors at the model level, we regress climate damages on an adjusted measurement of temperature change squared (following a common assumption in the literature that is supported by earlier meta-analyses) and its interaction with indicator variables for whether the model captures non-market, productivity, and catastrophic climate impacts. By dropping duplicate damage estimates and outliers, we explore the impact of duplicate publication and other biases on the damage-temperature relationship. In doing so, we are able to better characterize the current state of knowledge with respect to the damage-temperature relationship. To understand the potential effect of the previous bias on the SCC, we re-calibrate DICE – generally seen as the first and most respected IAM – by replacing its damage function with our preferred specification of the damage-temperature relationship and re-estimating the SCC.

Our paper is structured as follows. First, we discuss the relevant meta-analysis literature in the context of global climate damages. Second, we describe the creation of our datasets. Third, we present the econometric model used for our meta-analysis of willingness to pay. Fourth, we present our regression results with respect to the damage-temperature relationship. Fifth, we provide a short sensitivity analysis. Sixth, we discuss the interpretation of our results as a damage function. Finally, we conclude.

2. Conducting Meta-analyses of Global Climate Damage Estimates

Meta-analysis (and specifically meta-regression) has become a common tool in environmental economics. Nelson and Kennedy (2009) provide a thorough review of this literature up until 2009, and find a wide range of quality in the literature. Meta-regression is commonly used to study climate damages to a particular economic sector (e.g., Challinor et al. (2014) analyze global agricultural production) and/or region (e.g., Houser et al. (2014) analyze U.S. climate damages). Some of these meta-analyses reveal complex relationships between climate-change and society, including Hsiang et al. (2013) who find that higher temperatures and more extreme rainfall events – as predicted under most climate change scenarios – will lead to more human conflict. Only recently has meta-regression been employed at the macro-scale (Tol, 2009; Tol, 20013; Nordhaus and Sztorc, 2013; Newbold and Marten, 2014), though none have met the standards laid out in Nelson and Kennedy (2009).

In this study, we are interested in the factual and methodological determinants of *global willingness to pay to avoid total impacts of climate change* as measured as % of global GDP (*D*). Like previous meta-analyses, increase in global average surface temperature (*T*) is included as the sole factual cause of observed heterogeneity. The idea behind this decision is not that temperature is the singular climate-related driver of impacts, but that many of the other climate drivers of impacts, such as an increase in storm intensity and precipitation change, are strongly correlated with temperature change. Unlike previous meta-analyses of global climate damage estimates, we also control for methodological causes of the observed heterogeneity in the damage impact estimates to avoid omitted variable bias. We divide these methodological variables into those that directly (*R*) and indirectly (*W*) impact the relationship between damages-temperature.

Generally, the meta-analysis regression is

(1) $\widehat{D} = f(T, R, W) + \varepsilon$

where ε is the error term. This error term can be subdivided into unobservables (μ) and measurement errors (e) from studies where $\mu \sim N\left(0, \sigma_{\mu}^{2}\right)$ and $e \sim N(0, \sigma_{e}^{2})$.[1] While σ_{e}^{2} is often reported in statistical studies (Rhodes, 2012), this is unobserved for non-statistical (i.e., enumerative, CGE, and science-based) damage estimates. Following the common assumption of a linear function, we assume that damage-temperature relationship becomes

(2) $\widehat{D} = \gamma + g(T)\alpha + Rg(T)\beta + W\delta + \mu + e$ and $\mu = \alpha_1 Z + \alpha_2 X$

where *Z* and *X* are the unobserved factual and methodological causes of the unobserved heterogeneity of damages.

Heteroskedasticity and dependent errors are common issues in meta-analyses (Nelson and Kennedy, 2009; Stanely et al., 2013; EPA, 2006). Unaddressed (e.g., using OLS), these estimation issues result in inconsistent and inefficient coefficient estimates – which are particularly important given the small sample sizes of most meta-analyses – and biased estimates of standard errors. Heteroskedasticity arises because of significant unobserved heterogeneity in the damage estimates. Because the standard errors (i.e., σ_e^2) are unobserved for non-statistical global climate damage estimates, weighted least squares (WLS) estimators as traditionally specified (i.e., using $\frac{1}{\sigma_e^2}$ as weights) – the preferred estimator for small sample sizes – cannot be used,[2] and instead Davidson and MacKinnon or other heteroskedasticity robust standard errors are recommended (EPA, 2006; Nelson and Kennedy, 2009; Rhodes, 2012).

Dependence of the error terms arises because of correlations in the underlying damage estimate - including similar authorship, estimation method, or model - which result in clustering of error terms. Due to small sample size, the use of the preferred estimators to address dependence – generalized least squares (GLS) and panel estimators – is limited (Cameron and Trivedi, 2010, pages 274 to 275; Nelson and Kennedy, 2009; Bergh and Button, 1999), and a combination of cluster robust standard errors, carefully chosen control variables, and the use of only one estimate per study are recommended (Nelson and Kennedy, 2009; Cameron and Trivedi, 2010).

Outlier estimates is another key economic issue in meta-analysis due to their potential overly influential impact on study findings (Nelson and Kennedy, 2009). In the context of climate change, outlier damage estimates potentially correspond to temperature increases exceeding 3.2 °C (approximately the temperature increase from a doubling of atmospheric CO2 from the pre-industrial level). Above this temperature threshold, estimates are highly contentious and most of these estimates are potentially extrapolations of low damage estimates, violating the independence assumption. The meta-analysis literature recommends conducting sensitivity analysis to outliers by dropping potential outliers and using alternative outlier robust estimators (Huber weights, trimmed least squares, and minimum absolute deviation estimator) (EPA, 2006; Nelson and Kennedy, 2009).

3. Data

For our study's dataset, we combine the studies included in the most recent analyses by Tol (2013) and Newbold and Marten (2014). We double-check estimates – including re-aggregating damage estimates using GDP weights (Columbia, 2002) – and correct for any errors when present. We also searched for new damage estimates by: (1) including updates of previous damage estimates; (2) reviewing publications of already-included authors; (3) reviewing the damage estimates underlying the studies cited in the most up-to-date meta-analysis of the SCC (Havranek et al., 2014); and (4) searching Google Scholar and Econlit.[3] We include an estimate if it meets the following conditions: global, unique (i.e., not a re-running of a previously cited IAM or an IAM copying a previously cited IAM), and dating after 1993 (the start of modern climate damage estimates (Tol, 2009)). Following the EPA (2006) recommendation, we include only one estimate per study,[4] unless multiple estimates were deemed potentially independent. We assembled 49 damage estimates from 40 studies – using six estimation strategies (enumerative, statistical, CGE, survey, science-based, and experimental); 24 primary authors (including ten estimates from Nordhaus); and 22 to 27 models (depending on the definition). Of these estimates, 11 are drawn from the grey literature as means of addressing publication bias.[5]

After assembling the data, we drop estimates that do not meet our *a priori* selection criteria. We are interested solely in *global willingness to pay to avoid climate change*, and drop all studies that measure impacts in terms of compensating surplus: Maddison (2003), Rehdanz and Maddison (2005), and

Rehdanz and Maddison (2011).[6] Additionally, we exclude all estimates that arbitrarily cap damages based on author discretion: Nordhaus (1994a) and Nordhaus and Yang (1996).[7] Finally, we drop cross-sectional studies at the national level that do not address omitted variable bias: Choiniere and Horowitz (2006). Our *final* dataset includes 43 data points from 31 studies.

Given that many of the estimates are updates of or citations of previous estimates – giving a false sense of precision – we further limit our observations to prevent duplication bias (i.e., double counting) (Gotzsche, 1989; Tramer, 1997; Nelson and Kennedy, 2009). We define a study as a duplicate if it is not the most up-to-date estimate by an author utilizing a particular method or if the estimate cites already-included estimates. By dropping estimates that correspond to this definition, we construct a new dataset – the *non-cited* dataset – that consists of 26 non-duplicate estimates from 20 studies.

We code multiple damage, temperature, and methodological variables. *D_new* is the damage estimate cited in the reviewed paper. *T_new* is global average mean surface temperature increase in degrees Celsius relative to the pre-industrial or current time. For studies that estimate the impact of an increase in global land temperature, we multiply their temperature change by the ratio between global land and surface temperatures (as defined by NOAA's State of the Climate dataset) to convert to global surface temperature. We take care to distinguish studies that include non-market damages or catastrophic damages or allow for an effect through changes in productivity. Similarly we distinguish between cross-sectional and panel data and we account for the type and date of each study. Finally, we classify studies by author, estimation method, and model type.[8]

See Table 1 for a summary of the two datasets. We can clearly see that the mean damage estimate increases from 6.28% of GDP to 9.49% from the *final* to the *non-cited* datasets, respectively; a similar increase occurs for the median and temperature weighted mean damage estimates. Contrary to the findings of Tol (2009), these damage estimates appear to be increasing over time, though the results are generally insignificant.[9] If we look to Figure 2, we can see these damage estimates differ by author, estimation method, and model. In particular, damage estimates derived using science-based (particularly those from Weitzman) and survey methodologies tend to be higher, though there are a handful of high enumerative and statistical damage estimates. Furthermore, damage estimates derived by Nordhaus and Tol using the enumerative strategy appear to be lower than other enumerative estimates. Finally, looking at Figure 3, we can see that the bulk of estimates for both datasets are below 3.2 °C, highlighting the potential for high temperature estimates to have a disproportionate effect on meta-regression results.

4. Econometric Specification

Following historical precedent (Nordhaus and Sztorc, 2013, Newbold and Marten, 2014; Tol, 2014), we assume a quadratic damage function with no initial benefits from climate change such that $g(T) = T^2$ in equation *(2)*.[10] Unlike previous papers, we control for whether temperature is measured relative to pre-industrial or current temperatures by including an adjustment term θ_j – as measured by the difference in temperature in the base year (or period) of the study relative to the pre-industrial period[11] – such that equation (2) simplifies to

(3) $$D_{new\,j} = \alpha\left(T_j + \theta_j\right)^2 - \alpha\theta_j^{\,2} + \varepsilon_j = \alpha\left[T_j^{\,2} + 2\theta_j T_j\right] + \varepsilon_j = \alpha * t2_j + \varepsilon_j$$

where we assume that there are no climate damages for a 0 °C increase (a standard assumption implying $\gamma = 0$ in equation (3)) and $t2$ is our new temperature-squared variable that accounts for its corresponding estimate's base period. If we relax the implicit assumption in expression (3) that there are no methodological variables and account for clustering (i.e., dependence), we can rewrite this expression as

$$D_{new\,j,k} - \delta X_{j,k,Cross} = \alpha * t2_{j,k} + \beta_1 * t2_{j,k} * X_{j,k,mkt} + \beta_2 * t2_{j,k} * X_{j,k,cat} + \beta_3 * t2_{j,k} * X_{j,k,prod} + \varepsilon_{j,k}.$$

where *k* is the cluster level (*author, estimation method and model*) and $X_{j,k,l}$ is an indicator variable for estimate j and category $l \epsilon \{mkt, cat, prod, cross\}$ equal to one if estimate *j* captures only market impacts (i.e., omits non-market impacts), includes catastrophic impacts, accounts for productivity, and corresponds to a cross-sectional estimate, respectively. This expression simplifies to

(4) $$D_{new\,j,k} = \alpha t2_{j,k} + \beta_1 mkt\_t2_{j,k} + \beta_2 cat\_t2_{j,k} + \beta_3 t2_j prod\_t2_{j,k} + \delta X_{j,k,Cross} + \varepsilon_{j,k}$$

where $\varepsilon_{j,k} \sim N(0, \sigma^2)$.

Given our study's small sample size, we estimate expression (4) using WLS with clustered standard errors assuming we reject the null hypotheses of homoscedasticity and independent observations. In terms of addressing heteroskedasticity, we do not observe measurement error or the number of observations for many global climate damage estimates. Since the literature finds that damage estimates are more uncertain for higher temperature changes (Tol, 2009; Tol, 2014; Burke et al., 2015), we instead weight estimates by the inverse of temperature change to place lower weight on the more uncertain damage estimates (Day, 1999); this has the added benefit of placing lower weights on outlier estimates (i.e., estimates above 3.2 °C). In terms of addressing dependence of estimates, there are multiple scales of potential dependence: author, estimation method, and model. Therefore, we cluster at the scale for which there is the strongest empirical evidence of dependence, in addition to carefully choosing control variables and selecting only one estimate per study (except when potential independence cannot be rejected).

We estimate expression (4) with four datasets: *low-final*, *all-final*, *low-non-cited*, and *all-non-cited*. To demonstrate the importance of duplicate publication bias, we estimate expression (4) using the *final* and *non-cited* datasets that include and exclude duplicate estimates, respectively. Given that the *final* dataset is used only as a point of comparison to *non-cited,* it should be assumed that we are referring to the *non-cited* datasets unless explicitly stated. To address outliers,[12] we estimate expression (4) using estimates corresponding to low temperature increases (i.e., 3.2 °C or below) and data corresponding to all temperature increases. To signify the different treatment of outliers, we include the additional prefix *low* and *all* to the dataset name. For purposes of simplicity, we will also refer to the *low-non-cited* and *all-non-cited* datasets as the *low* and *all* datasets, respectively, given the focus on the *non-cited* results. Given the four potential datasets developed for this paper, our preferred dataset is the *low* (e.g., the *low-non-cited*) dataset for analyzing the potential impact of climate change in the next century.

Finally, given concerns about overfitting our model due to its small sample size – particularly in our preferred dataset consisting of 21 observations – we conduct our analysis using two sets of exogenous variables: a smaller set that includes only temperature squared and its interaction with the market and catastrophic indicator variables and an extended set that also includes the indicator variable for cross-sectional data at the country scale and the interaction of temperature squared with the productivity indicator variable. The latter two variables are chosen for exclusion based on their relatively lower variation in the preferred (i.e., the *low*) dataset.[13]

5. Results

Using the *low* and *all* datasets that exclude and include outliers, we test for heteroskedasticity and dependent errors terms using Cameron & Trivedi's decomposition of IM-test (see Table 2) and Breusch-Pagan test of independence (see Table 3), respectively. We consistently reject the null hypothesis of homoscedasticity for the *all* dataset and consistently fail to reject the null hypothesis for the *low* dataset.[14] For purposes of simplicity, we present results using WLS for both datasets; these results only slightly differ from the OLS results when considering the *low* dataset. We consistently reject the null of independence at the *model* scale at the 1% significance level, and less consistently at the *author* and *method* scales at the 5% and 10% levels. For all datasets, we cluster error terms at the *model* scale.

Regressing climate damages on temperature squared and its interaction with indicator variables, the results indicate a strong relationship between temperature and climate damages; see Table 4. In particular, the regressions are all jointly significant at the 1% significant levels, and the coefficient corresponding to temperature-squared (i.e., non-catastrophic climate damages) is positive and significant at the 5% significance level. Additionally, catastrophic impacts are positive, but only significant at the 1% level when damage estimates for high temperature increases (i.e., above 3.2 °C) are included. As expected, we generally find that the exclusion of non-market impacts has a negative and sometimes significant effect on damages. In the extended specifications, we find that productivity has an insignificant, positive effect on climate damages and strong evidence of cross-sectional bias.

Our results appear split with respect to our other two main statistical concerns – over-specification and outliers. On the one hand, over-specification appears to not be of particular concern given only the small change in coefficients corresponding to temperature-squared and catastrophic impacts that occurs with the inclusion of the productivity and cross-sectional controls. While there are some significant changes in the coefficient corresponding to the market indicator variable, this is due to the high correlation between the market and productivity indicator variables in the *low* (0.47%) and *all* (57%) datasets that results from all *CGE* models excluding non-market impacts and modeling productivity. On the other hand, outlier estimates appear to be a critical issue given the large change in coefficient magnitudes between the *low* and *all* datasets, though only the coefficient corresponding to temperature-squared and its interaction with the market indicator variable are statistically different.[15] The signs of coefficients are almost identical between the *low* and *all* datasets; the only difference again is the coefficient corresponding to the market indicator variable for the *all* dataset due to the high level of correlation between the market variable and productivity. As a consequence of this correlation between the market and productivity impacts of climate change in the underlying damage estimates and an increase in adjusted-R squared with the inclusion of productivity and cross-sectional controls, we choose the extended specification as the preferred specification for both the *low* and *all* datasets.

Our damage estimates are statistically higher than Nordhaus and Sztorc (2013). We reject the null hypothesis that non-catastrophic damages (the coefficient corresponding to *t2*) and total damages (the sum of the coefficients corresponding to temperature-squared and its interaction with the catastrophic dummy) are equivalent to the Nordhaus and Sztorc (2013) damage estimate of 0.2132. If we also consider climate change's impact on productivity jointly (i.e., add the coefficient corresponding to the interaction of temperature-squared and productivity) with non-catastrophic and total impacts, this difference remains statistically significant.[16]

In support of the main hypothesis of our paper, we find strong evidence of duplication bias for our preferred dataset that excludes damage estimates above 3.2 °C (i.e., the *low* dataset). For this dataset, the damage coefficient corresponding to temperature-squared increases (by approximately 80% for the preferred specification) and the coefficient corresponding to catastrophic impacts declines as duplicate estimates are dropped. As expected, the standard errors corresponding to these variables also increase as more data is dropped – resulting in catastrophic impacts becoming statistically insignificant – emphasizing the false precision resulting from duplication. Though less substantial, we find similar changes for the larger dataset that includes outlier estimates (i.e., the *all* dataset). Using seemingly unrelated estimation to test for duplicate publication bias (i.e., a significant difference in damage estimates between the *final* and *non-cited* datasets), we reject the null of no duplication bias at the 1% significance level for our preferred dataset that considers only damage estimates corresponding to low temperature increase (i.e., 3.2 °C or less) and fail to reject the null for the dataset containing potential outliers.[17] Interestingly, while we find strong evidence of duplication bias, we find no evidence of publication bias.[18]

For our preferred regression (i.e., regression (4) in Table 4 corresponding to the regression of damages on the extended set of variables using the *low* dataset), we have shown that the coefficient corresponding to temperature-squared – our current best estimate of non-catastrophic impacts of climate change – significantly increases (almost doubling) when we address duplication bias and that the resulting estimates differ from Nordhaus and Sztorc (2013). Given that our results demonstrate that duplication bias can significantly affect the temperature-damage relationship, it is important to determine the relative importance of duplication bias in driving the difference in our results from Nordhaus and Sztorc (2013) relative to the other forms of bias discussed earlier – omitted variable bias, publication bias, and measurement error – and relative to obtaining a larger dataset and improving the estimation technique.

To address this issue, we run the Nordhaus and Sztorc (2013) model over eight datasets and sets of variables to identify the relative impact of each adjustment made from the Nordhaus and Sztorc (2013) estimate – specification (1) – to our preferred estimates - specification (8); see Table 5. Each step captures the impact of one particular change that we introduced: (2) improving the estimation strategy (including data corrections), (3) introducing additional published damage estimates, (4) introducing grey literature (i.e., accounting for publication bias), (5) correcting the temperature specification for the base period, (6) accounting for omitted variable bias, (7) accounting for duplication bias, and (8) introducing the extended set of control variables. Considering non-catastrophic impacts and productivity jointly, we find that duplication bias has the most significant impact on the % change in the coefficient corresponding to temperature-squared (81%), followed by introducing new published data (54%) and introducing productivity and cross-sectional controls (31%). If we consider the percentage change in total climate damages (non-catastrophic plus catastrophic) and productivity jointly, omitted variable bias (91%) is the most important source of bias followed by new published data (54%) and then duplication bias (24%).

If we make the assumption that the relationship captured in our meta-analysis represents a damage function (Tol, 2009; Nordhaus and Sztorc, 2013), we may be interested in understanding which factors – including source of bias – have the most influence on the SCC. Using DICE – the sole IAM to currently calibrate its damage function using meta-analysis and one of three models used to calculate the official U.S. SCC (IWG, 2013) – we estimate the SCC and repeat the above analysis. If we consider only non-catastrophic impacts (as in the DICE-2013R damage function), addressing the above biases increases the SCC (in 2015) by 203% to 314% from specification (1) to (8) depending on whether we account for the impact of climate change on economic productivity. If we consider total climate impacts (both non-catastrophic and catastrophic), addressing the above biases increases the SCC (in 2015) by 344% to 469% depending on the treatment of productivity; see Figure 4. When considering the % change in the SCC from specification to specification, we find identical results to the previous paragraph indicating that duplication is one of the most significant sources of bias in previous estimates. Finally, we find that uncertainty over the SCC increases – as measured by the difference between the 5th and 95th percentile SCC estimates (accounting for non-catastrophic impacts only) – from specification (1) to (8).

Unsurprisingly given the insignificance of duplicate publication bias found earlier when considering data corresponding to all temperature increases, we find that accounting for duplication bias has a similarly small effect on the temperature-damage relationship and the SCC when considering outliers; see Table 6. However, the 2015 SCC estimates corresponding to this dataset are still far above that captured in the original DICE-2013R model, with an increase in the range of 52% to 320% when considering only non-catastrophic climate impacts and 211% to 521% when considering total climate impacts where the range is dependent on the inclusion of the impact of climate change on productivity. When we consider the climate impacts on productivity jointly with the non-catastrophic and total impacts of climate change, our final SCC estimates – specification (8) – are almost identical regardless of whether we account for outliers, and considerably higher and more uncertain than the original DICE-2013R estimate.

6. Sensitivity Analysis

Additional sensitivity analyses were run for the preferred specifications in Table 4 excluding and including the outliers (i.e., regressions (4) and (8) in Table 4, respectively); see the *Additional Material*. First, we re-run our results using Tol's (2009) functional form – which includes an additional linear temperature term; see Tables A3 to A5. We find strong evidence that the results differ from Tol (2009), and even stronger evidence of duplication bias.[19] Second, we analyze the impact of a more restrictive definition of duplication (see Table 7); this result is discussed further in the following section. Third, we re-run the preferred specification with clustering at three alternative scales: author, estimation method, and an alternative definition of model; see Table A6. The results are only slightly less significant. Fourth, we re-estimate the preferred specification using alternative estimators: OLS, GLS, and panel fixed effects at the model scale; see Table A7. Not only are our results generally robust to these alternative estimators, the non-catastrophic and total climate damages implied by our preferred estimator – weighted least squares – are, if anything, lower bounds. Fifth, we run a variety of alternative specifications, and find our results are fairly robust (particularly the coefficient corresponding to *t2*); see Tables A8 and A9.[20] Last, we re-run the preferred regression dropping each of the observations (see Figures A6 and A7), and re-estimate the model using outlier robust estimators (see Table A10). We find that the results are fairly robust to this exercise.[21]

7. Discussion

Like Nordhaus and Sztorc (2013), we take the extra step of interpreting the temperature-damage relationship captured by the above meta-analysis as a climate damage function. To include these meta-analysis results in DICE-2013R, we followed Nordhaus in multiplying the sum of the coefficients corresponding to non-catastrophic impacts (temperature-squared and potentially its interaction with productivity) by 1.25 to account for omitted non-catastrophic climate impact.[22] Given our more improved estimation strategy using multiple datasets (*low* and *all*) and sets of control variables, it is important to clarify which of our estimates is the appropriate relationship to use in replacing the DICE-2013R damage function.

When re-estimating the DICE-2013R damage function, excluding outliers (i.e., choosing the *low* dataset that consists of only damage estimates corresponding to temperature increases of 3.2 °C or less) is appropriate on several grounds. First, outlier estimates are highly speculative in nature given that uncertainty in climate damage estimates increases in temperature, and yet including these five estimates (19% of estimates) significantly decreases the impact of temperature-squared on non-catastrophic damages by half and shifts catastrophic impacts from statistically insignificance to significance at the 1% level (see Table 4); this is despite the inclusion of weights to partially mitigate this undue influence. In contrast to the outsized impact of these estimates, the majority (approximately two-thirds) of the 2015 SCC estimates for DICE-2013R correspond to impacts occurring this century; this result is relatively robust to the magnitude of the damage coefficient. Thus, excluding damage estimates for temperature changes above 3.2 °C prevents outlier estimates from unduly influencing the estimated temperature-damage relevant for the next century for which estimates for approximately 3.2 or less are more germane.

Even more problematic than the undue influence of the outlier estimates, the majority of damage estimates corresponding to high temperature increases (greater than 3.2 °C) are drawn from studies from which another damage estimate is already included. While we assume that these estimates are independent in our base analysis, it is possible that one estimate is an extrapolation of the other leading to duplicate publication bias. If we adopt a more restrict definition of duplication that allows for only one estimate per study in addition to the previous definition – which we call the *unique* dataset[23] – the results corresponding to the low temperature estimates are robust while the estimates corresponding to all temperature estimates are not; see Table 7.[24] Thus, the results corresponding to the *low* dataset are also more robust to the definition of duplication, making them more stable for use as the basis of a damage function.

Following Nordhaus and Sztorc (2013), we specify a damage function corresponding to non-catastrophic damages from regression (4) in Table 4 excluding productivity such that the quadratic damage function parameter is equivalent to the coefficient corresponding to temperature-squared. Thus, while we control for catastrophic damages and productivity in our preferred regression specification, we exclude the corresponding estimated coefficients in the base specification of the damage function because of their mixed significance and volatility across the various specifications. Instead, given the debate over the impact of climate change on productivity and economic growth (Dell et al., 2012; Burke et al., 2015; Howard and Sylvan, 2015), we recommend conducting a sensitivity analysis to the inclusion of the productivity impact. To account for catastrophic impacts, we recommend running a Monte Carlo simulation whereby catastrophic impacts are captured by the variance of the damage coefficient (hence the importance of improved estimation of standard errors). Replacing the DICE-2013R damage function with our estimates of non-catastrophic damages, we find a mean 2015 SCC estimate of $68 (2015 USD) with a 95% confidence interval from $16 to $132 using our base specification and a mean of $93 with a range of -$2 to $234 when including productivity.

Given that DICE is often run as a deterministic model, we recommend also conducting a sensitivity analysis with respect to the inclusion of catastrophic damages if a Monte Carlo simulation is not run. Replacing the DICE-2013R damage function with our estimate of total damages (non-catastrophic and catastrophic), we find a mean 2015 SCC estimate of $100 (2015 USD) and $128 when we exclude and include productivity, respectively.

Some have questioned if this type of top-down analysis is appropriate, and thus question the validity of SCC estimates such as those derived above. For example, Tol (2009) argues that these studies should not be treated as time-series data, and that any analysis attempting to estimate a damage function should be interpreted cautiously. Citing the IPCC (2014), JEP (2015) argues against this methodology because estimates use different estimation methods and models to capture diffing overlapping sets of impacts. In this paper, we attempt to address these short-comings by including methodological variables that capture differences in overlapping estimates and accounting for different estimation methods and models. While this type of meta-analysis is fraught with challenges as critics suggest, we believe that analysts should continue to work to improve these estimates because such analyses provide valuable information about the temperature-damage relationship and some analysts have already chosen to embrace the technique (Nordhaus and Sztorc, 2013); though, analysts should also be clear about the limitations of the methodology. Furthermore, as more data becomes available, we should expect meta-analysis estimates to improve as analysts are able to control for a more refined set of differences between estimates (including which sectors are included such as agriculture, sea-level rise, etc.).

As of late, there has been a new wave of advanced statistical estimates of climate damages – including bottom-up estimates (Schlenker and Roberts, 2009; Hsiang et al., 2011) and global market impact estimates (e.g., Dell et al;, 2012 Burke et al., 2015) – that some economists may argue make the top-down approach to estimating a damage function unnecessary. However, the above techniques are also key to the bottom-up approach given that most of these studies focus on particular sectors in developed nations only, and will require benefit-cost transfer and meta-analyses methods – as developed in this paper – to assemble global climate damage estimates. For example, there are a multitude of regional agricultural studies on the costs of climate change, and meta-analyses are necessary to estimate regional and global agricultural damage functions.[25] In the meantime, as statistical methods and data improve, the top-down approach – meta-analysis and surveys – are necessary. Climate policy cannot wait while statistical studies are perfected. Instead, we should utilize the best available estimates of total climate damages, and update these estimates as new data becomes available. We should expect that over time – as both methods improve and more data become available – top-down and bottom-up approaches should converge. Future work should aim to improve global climate damage estimates (the top-down approach) and region-sector climate damage estimates (a bottom-up approach), and utilize meta-analyses to determine whether these approaches are converging in magnitude.

8. Conclusion

This paper hopes to make contributions to the literature on IAMs and climate change more generally, by offering an improved estimation of the damage–temperature relationship. Using a larger dataset than previously assembled, this paper highlights several methodological improvements that address bias present in previous meta-analyses. Specifically, due to restricted analyses resulting from a limited availability of data, previous meta-analyses of climate damages suffered from multiple sources of bias: duplication bias, measurement error, and omitted variable bias. By expanding our dataset (to include additional published and grey literature estimates), including methodological variables, and correcting the specification of temperature (to account for different reference periods), we are able to address and test for these biases. Estimating the relationship between temperature and climate damages using cluster robust standard errors at the model level – to address dependence of observations – we demonstrate that duplication bias and omitted variable bias are likely present and have the most significant impact on the results. As a consequence, the temperature-damage relationship estimated in Nordhaus and Sztorc (2013) is biased downwards by significant amounts with the exact amount of the bias depending on how we account for climate impacts on economic productivity and potential catastrophic impacts of climate change. We also demonstrate that decisions regarding the treatment of outliers (i.e., damage estimates corresponding to temperature increases exceed 3.2 °C) are critical in determining this underlying relationship. In doing so, we highlight the need to carefully consider the relevant time period over which the relationship need be identified.

To the authors' knowledge, this paper introduces the concept of duplicate publication from the medical literature (Tramer et al., 1997; Gotzsche, 1989) into the climate economics literature. Unlike the medical literature that discusses the potential bias from overt and covert re-publication of scientific results, we are concerned with potential bias from the common practices of updating climate-damage estimates over time and calibrating climate-model damage functions on previous estimates in the climate damage literature. As a consequence, many previous meta-analyses of global climate damage estimates contain multiple damage estimates from practically identical models. If we exclude the impact of outlier estimates (i.e., damage estimates corresponding to a greater than 3.2 °C temperature increase), we find strong evidence that duplication bias has a sizable and statistically significant impact on the relationship between temperature and climate damages (as a % of GDP).

We can interpret the damage-temperature relationship estimated above as a damage function (as done in DICE-2013R). Ignoring outlier damage estimates (i.e., damage estimate corresponding to temperature increases above 3.2 °C) and replacing the DICE-2013R damage function – which suffered from the above biases – we find that the 2015 SCC increases by approximately 300% to 400% depending on the treatment of productivity. Though DICE-2013R has yet to be integrated into the U.S. government's analysis (most recently the Interagency Working Group on the Social Cost of Carbon (IWG) cited DICE-

2010 which relies on the previously employed enumerative calibration strategy), this paper highlights how the U.S. government could re-estimate the DICE-2013R damage function to meet previously set government standards for meta-analyses (EPA, 2006).

References

van den Bergh, J. C. J. M., & Button, K. (1999). Meta-analysis, economic valuation and environmental economics. *Handbook of Environmental and Resource Economics*, 796-808.

Burke, M., Hsiang, S. M., & Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature*.

Cameron, C. A. and Trivedi, P., K. (2010). *Microeconometrics using Stata: Revised Edition*.

Challinor, A. J., Watson, J., Lobell, D. B., Howden, S. M., Smith, D. R., & Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change*, *4*, 287-291.

Choiniere, C.J., and Horowitz, J.K. (2006). A Production Function Approach to the GDP-Temperature Relationship. March. *Preliminary version of the paper with temperature data from Worldclimate.org.*

[Columbia, 2002] Center for International Earth Science Information Network (CIESIN). (2002). Country-level GDP and Downscaled Projections based on the A1, A2, B1, and B2 Marker Scenarios, 1990-2100, [digital version]. Palisades, NY: CIESIN, Columbia University. Available at http://www.ciesin.columbia.edu/datasets/downscaled. Available January 28, 2016.

Day, B. (1999). A meta-analysis of wage-risk estimates of the value of statistical life. In: EC, Benefits transfer and the economic valuation of environmental damages, European Commission, Brussels, pp. 1-29.

Dell, M., Jones, B. F., & Olken, B. A. (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics*, 66-95.

[EPA, 2006] U.S. Environmental Protection Agency. (2006). Report of the EPA Working Group on VSL Meta-analysis. Available at http://yosemite.epa.gov/ee/epa/eerm.nsf/vwAN/EE-0494-01.pdf/$file/EE-0494-01.pdf.

Gøtzsche, P. C. (1989). Multiple publication of reports of drug trials. *European journal of clinical pharmacology*, *36*(5), 429-432.

Havranek, T., Irsova, Z., Janda, K., & Zilberman, D. (2014). Selective reporting and the social cost of carbon.

Houser, T., Hsiang, S., Kopp, R., Larsen, K., Delgado, M., Jina, A., ... & Rising, J. (2014). American climate prospectus. *Economic Risks in the United States. Rhodium Group, LLC*.

Hsiang, S. M., Burke, M., & Miguel, E. (2013). Quantifying the influence of climate on human conflict. *Science*, *341*(6151), 1235367.

Howard, P. H. (2014). Omitted Damages: What's Missing From the Social Cost of Carbon. *The Cost of Carbon Pollution Project.* Available at http://costofcarbon.org/reports/entry/omitted-damages-whats-missing-from-the-social-cost-of-carbon.

Howard, P. H., & Sylvan, D. (2015). The Economic Climate: Establishing Expert Consensus on the Economics of Climate Change. *Institute for Policy Integrity* Working Paper No. 2015/1.

Hsiang, S. M., Meng, K. C., & Cane, M. A. (2011). Civil conflicts are associated with the global climate. *Nature*, *476*(7361), 438-441.

[IPCC, 2013] Intergovernmental Panel on Climate Change. (2013). The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge, United Kingdom and New York, USA*.

[IPCC, 2014] Intergovernmental Panel on Climate Change. Working Group II. (2014).*Climate Change 2014: Impacts, Adaptation, and Vulnerability*.

[IWG, 2013 Revised] Interagency Working Group on Social Cost of Carbon. (2013). Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866. United States Government. Available at http://www.whitehouse.gov/sites/default/files/omb/assets/inforeg/technical-update-social-cost-of-carbon-for-regulator-impact-analysis.pdf.

[JEP, 2015] Editors of Journal of Economic Perspectives. (2015). Editorial Note: Correction to Richard S. Tol's "The Economic Effects of Climate Change." *Journal of Economic Perspectives,* 29(1), 217-220.

Kopp, R. E., Hsiang, S. M., & Oppenheimer, M. (2013). Empirically calibrating damage functions and considering stochasticity when integrated assessment models are used as decision tools.

Maddison, D. (2003). The amenity value of the climate: the household production function approach. *Resource and Energy Economics*, *25*(2), 155-175.

Maddison, D., & Rehdanz, K. (2011). The impact of climate on life satisfaction. *Ecological Economics*, *70*(12), 2437-2445.

Nelson, J. P., & Kennedy, P. E. (2009). The use (and abuse) of meta-analysis in environmental and natural resource economics: an assessment. *Environmental and Resource Economics*, *42*(3), 345-377.

Newbold, S. C., & Marten, A. L. (2014). The value of information for integrated assessment models of climate change. *Journal of Environmental Economics and Management*, *68*(1), 111-123.

[Nordhaus, 1994a] Nordhaus, W. D. (1994). *Managing the global commons: the economics of climate change* (Vol. 31). Cambridge, MA: MIT press.

[Nordhaus, 1994b] Nordhaus, William D. (1994). "Expert Opinion on Climate Change." *American Scientist,* 82(1):45–51.

Nordhaus, W. D. (2006). Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3510-3517.

Nordhaus, W. (2008). New metrics for environmental economics: Gridded economic data. *Integrated Assessment*, *8*(1).

Nordhaus, W. D., & Boyer, J. (2000). *Warning the World: Economic Models of Global Warming*. MIT Press (MA).

Nordhaus, W. D., & Sztorc, P. (2013). *Introduction and User's Manual. Available at* http://www.econ.yale.edu/~nordhaus/homepage/documents/Dicemanualfull.pdf.

Nordhaus, W. D., & Yang, Z. (1996). A regional dynamic general-equilibrium model of alternative climate change strategies. *The American Economic Review*, 741-765.

Revesz, R. L., Howard, P. H., Arrow, K., Goulder, L. H., Kopp, R. E., Livermore, M. A., ... & Sterner, T. (2014). Global warming: Improve economic models of climate change. *Nature*, *508*(7495), 173-175.

Rhodes, W. (2012). Meta-Analysis An Introduction Using Regression Models. *Evaluation review*, *36*(1), 24-71.

Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to US crop yields under climate change. *Proceedings of the National Academy of sciences*, *106*(37), 15594-15598.

Sterner, T. 2015. "Higher costs of climate change". Nature, 527:177-178. DOI:10.1038/nature15643

Tol, R. S. J. (2009). The economic effects of climate change. *The Journal of Economic Perspectives*, *23*(2), 29-51.

Tol, R. S. (2013). *Bootstraps for Meta-Analysis with an Application to the Impact of Climate Change* (No. 6413).

Tol, R. S. (2014). Correction and Update: The Economic Effects of Climate Change. *The Journal of Economic Perspectives*, 221-225.

Tramèr, M. R., Reynolds, D. J. M., Moore, R. A., & McQuay, H. J. (1997). Impact of covert duplicate publication on meta-analysis: a case study. *Bmj*, *315*(7109), 635-640.

Tables

Table 1a. Summary of "final" dataset

Variable	Obs	Mean	Std. Dev.	Min	Max	Predict
D_new	43	6.289079	16.731	-2.3	99	NA
T_new	43	2.788023	1.846353	0.69	12	-
T2_new	43	11.10281	22.12024	0.4761	144	+
cat	43	0.209302	0.411625	0	1	-
cross	43	0.046512	0.213083	0	1	?
current	43	0.488372	0.505781	0	1	+
Eco_Market	43	0.116279	0.324353	0	1	?
Grey	43	0.232558	0.427463	0	1	?
Market	43	0.441861	0.502486	0	1	-
prod	43	0.255814	0.441481	0	1	+
Time	43	12.74419	7.098266	0	21	+

Table 1b. Summary of "non-cited" dataset

Variable	Obs	Mean	Std. Dev.	Min	Max	Predict
D_new	26	9.491381	21.01565	-1.42	99	NA
T_new	26	3.059231	2.290869	0.69	12	-
T2_new	26	14.40512	28.03787	0.4761	144	+
cat	26	0.269231	0.452344	0	1	-
cross	26	0.076923	0.271747	0	1	?
current	26	0.423077	0.503832	0	1	+
Eco_Market	26	0.038462	0.196116	0	1	?
Grey	26	0.192308	0.401919	0	1	?
Market	26	0.423077	0.503832	0	1	-
prod	26	0.192308	0.401919	0	1	+
Time	26	13.38462	7.483726	0	21	+

Table 2. P-values of Cameron & Trivedi's decomposition of IM-test for Preferred Specification using D_new, by dataset

Dataset	Low		All	
	final	non-cited	final	non-cited
Shorter variable set				
Heteroskedasticity	0.3783	0.118	0	0.0015
Skewness	0.0185	0.1242	0.1467	0.298
Kurtosis	0.1505	0.4353	0.1499	0.1683
Total	0.0473	0.0856	0	0.0026
Extended variable set				
Heteroskedasticity	0.6401	0.2957	0.0003	0.0114
Skewness	0.0556	0.3275	0.345	0.6171
Kurtosis	0.1983	0.4145	0.1513	0.1766
Total	0.2055	0.3069	0.0007	0.03

*Shorter variable set include temperature-squared (*t2)* and its interaction with market (*mkt_t2*) and catastrophic (*cat_t2*) dummies, while the extended variable set also includes the interaction of temperature-squared with an indicator variable for productivity (*prod_t2*) and an indicator variable for estimation with a cross-sectional dataset at the country scale (*cross*)
***The Low* dataset includes climate damage estimates corresponding to temperature increases of 3.2 °C or less (i.e., excludes outliers), while the *all* dataset includes damage estimates corresponding to all temperature increases (i.e., includes outliers). The *non-cited* dataset includes only the most up-to-date damage estimates that are not citations (i.e., not based on previously included estimates), while the *final* dataset fails to account for duplication in any way.

Table 3. P-values of Breusch-Pagan LM Test of Independence, by cluster-level and dataset

Dataset / Cluster Level	Low		All	
	final	non-cited	final	non-cited
author	0.0013	0.0620	0.0001	0.0119
method	0.0293	0.0293	0.2296	0.1689
model	0.0000	0.0003	0.0000	0.0000

**The Low* dataset includes climate damage estimates corresponding to temperature increases of 3.2 °C or less (i.e., excludes outliers), while the *all* dataset includes damage estimates corresponding to all temperature increases (i.e., includes outliers). The *non-cited* dataset includes only the most up-to-date damage estimates that are not citations (i.e., not based on previously included estimates), while the *final* dataset fails to account for duplication in any way.

Table 4. Base Regressions: WLS with Cluster Robust Standard Errors at the Model Level, by dataset

Dataset	*Low* : Damages for Temp. Increases<3.2 °C				*All*: Damages for All Temp. Increases			
	final		*non-cited*		*final*		*non-cited*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	D_new	D_new	D_new	D_new	D_new	D_new	D_new	D_new
t2	0.329**	0.329**	0.595**	0.595**	0.264***	0.264***	0.318***	0.318**
	(0.121)	(0.125)	(0.180)	(0.191)	(0.0590)	(0.0606)	(0.0976)	(0.102)
mkt_t2	-0.155	-0.0856	-0.477**	-0.680**	0.0284	-0.0204	0.0444	-0.403**
	(0.138)	(0.276)	(0.182)	(0.244)	(0.153)	(0.250)	(0.218)	(0.179)
cat_t2	0.422*	0.422*	0.288	0.288	0.413***	0.413***	0.364***	0.364***
	(0.207)	(0.213)	(0.276)	(0.293)	(0.0600)	(0.0616)	(0.0988)	(0.103)
prod_t2		-0.139		0.183		0.0401		0.471
		(0.260)		(0.153)		(0.292)		(0.271)
cross		2.452**		3.162***		2.452***		3.162***
		(0.824)		(0.915)		(0.812)		(0.883)
Observations	37	37	21	21	43	43	26	26
R2	0.532	0.594	0.676	0.759	0.841	0.847	0.862	0.873
Adjusted R-squared	0.491	0.531	0.621	0.683	0.829	0.827	0.843	0.843
Liklihood	-83.13	-80.49	-46.41	-43.30	-114.4	-113.5	-73.07	-71.96
F-statistic	17.01	23.19	19.78	15.61	1879	1110	1040	572.3
Prob>F	0.000192	1.69e-05	0.000466	0.000601	0	0	0	0
Hypothesis: non-catastrophic impacts (captured t2) equal Nordhaus and Sztorc (2013)'s estimate of 0.2132								
p-value	0.3569	0.3709	0.0665	0.0803	0.4023	0.4140	0.3082	0.3291
Hypothesis: total impacts (captured by t2 + cat_t2) equal Nordhaus and Sztorc (2013)'s estimate of 0.2132								
p-value	0.0057	0.0069	0.0053	0.0072	0.0000	0.0000	0.0000	0.0000
Hypothesis: non-catastrophic impacts (captured t2 + prod_t2) equal Nordhaus and Sztorc (2013)'s estimate of 0.2132								
p-value	.	0.9381	.	0.0499	.	0.7645	.	0.0752
Hypothesis: total impacts (captured by t2 + prod_t2 + cat_t2) equal Nordhaus and Sztorc (2013)'s estimate of 0.2132								
p-value	.	0.2705	.	0.0078	.	0.1095	.	0.0062

Robust standard errors in parentheses

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

** t2* is adjusted temperature squared; *mkt_t2* is adjusted temperature squared interacted with *market*; *cat_t2* is adjusted temperature squared interacted with *cat; prod_t2* is adjusted temperature squared interacted with *prod*, and *cross* is an indicator variable for country cross-sectional data.

**** The *non-cited* dataset includes only the most up-to-date damage original estimates, while the *final* dataset fails to account for duplication in any way.

Table 5. Potential Bias: OLS and WLS with Cluster Robust Standard Errors at the Model Level and the Corresponding 2015 SCC Estimate using the Resulting Estimates as a DICE Damage Function, using multiple specifications and datasets corresponding to low temperature increases only (3.2 °C or lower)

Dataset	*Low* : Damages for Temp. Increases<3.2 °C								
	Nordhaus and Sztorc (2013)	Originial	Correct Data	New Data - Published	New Data - All	Correct Temp	Omitted Variables	Duplication Bias	Omitted Variables 2
	-	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	-	D_orig	D_new	D_new	D_new	D_new	D_new	D_new	D_new
T2	0.2136	0.213***							
	-	(0.0640)							
T2_new	-		0.238***	0.366***	0.428***				
	-		(0.0498)	(0.0975)	(0.0863)				
t2	-					0.349***	0.329**	0.595**	0.595**
	-					(0.0870)	(0.121)	(0.180)	(0.191)
mkt_t2	-						-0.155	-0.477**	-0.680**
	-						(0.138)	(0.182)	(0.244)
cat_t2	-						0.422*	0.288	0.288
	-						(0.207)	(0.276)	(0.293)
prod_t2	-								0.183
	-								(0.153)
cross	-								3.162***
	-								(0.915)
Observations	-	13	9	28	37	37	37	21	21
R2	-	0.480	0.460	0.389	0.429	0.385	0.532	0.676	0.759
Adjusted R-squared	-	0.437	0.393	0.367	0.414	0.368	0.491	0.621	0.683
Liklihood	-	-22.79	-16.02	-63.48	-86.81	-88.20	-83.13	-46.41	-43.30
F-statistic	-	11.08	22.85	14.09	24.59	16.09	17.01	19.78	15.61
Prob>F	-	0.00601	0.00306	0.00376	0.000430	0.00204	0.000192	0.000466	0.000601
Hypothesis: non-catastrophic impacts equal Nordhaus and Sztorc (2013)'s estimate of 0.002132									
p-value	-	-	0.6364	0.1478	0.0300	0.1468	0.3569	0.0665	0.0499
Hypothesis: total impacts equal Nordhaus and Sztorc (2013)'s estimate of 0.002132									
p-value	-	-	-	-	-	-	0.0057	0.0053	0.0078
2015 SCC (2015 USD per metric ton of CO2e)									
Non-cat (5%)	-	$8	$12	$15	$25	$16	$6	$19	-$2
Non-cat (50%)	$22	$22	$25	$40	$47	$38	$35	$68	$93
Non-cat (95%)	-	$38	$39	$67	$71	$61	$68	$128	$234
Total (50%)	-	-	-	-	-	-	$78	$100	$128

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

* *T2* and *D_orig* are the temperature-squared and damage variables cited in Nordhaus and Sztorc (2013) squared; *T2_new* and *D_new* are these variables corrected for citation error only. See Table 4 for remaining notation.

** Specifications: (1) Nordhaus and Sztorc (2013), (2) improving the estimation strategy (including data corrections), (3) introducing additional published damage estimates, (4) introducing grey literature (i.e., accounting for publication bias), (5) correcting the temperature specification for the base period, (6) accounting for omitted variable bias, (7) accounting for duplication bias, and (8) introducing the extended set of control variables (our preferred specification).

Table 6. Potential Bias: OLS and WLS with Cluster Robust Standard Errors at the Model Level and the Corresponding 2015 SCC Estimate using the Resulting Estimates as a DICE Damage Function, using multiple specifications and datasets corresponding to all temperature increases

Dataset	*All*: Damages for All Temp. Increases								
	Nordhaus and Sztorc (2013)	Originial	Correct Data	New Data - Published	New Data - All	Correct Temp	Omitted Variables	Duplication Bias	Omitted Variables 2
	-	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	-	D_orig	D_new	D_new	D_new	D_new	D_new	D_new	D_new
T2	0.002136	0.213***							
	-	(0.0640)							
T2_new	-		0.238***	0.654***	0.640***				
	-		(0.0498)	(0.0972)	(0.0995)				
t2	-					0.554***	0.264***	0.318***	0.318**
	-					(0.0933)	(0.0590)	(0.0976)	(0.102)
mkt_t2	-						0.0284	0.0444	-0.403**
	-						(0.153)	(0.218)	(0.179)
cat_t2	-						0.413***	0.364***	0.364***
	-						(0.0600)	(0.0988)	(0.103)
prod_t2	-								0.471
	-								(0.271)
cross	-								3.162***
	-								(0.883)
Observations	-	13	9	33	43	43	43	26	26
R2	-	0.480	0.460	0.780	0.756	0.758	0.841	0.862	0.873
Adjusted R-squared	-	0.437	0.393	0.773	0.750	0.752	0.829	0.843	0.843
Liklihood	-	-22.79	-16.02	-96.98	-123.6	-123.4	-114.4	-73.07	-71.96
F-statistic	-	11.08	22.85	45.30	41.33	35.31	1879	1040	572.3
Prob>F	-	0.00601	0.00306	2.10e-05	2.24e-05	4.88e-05	0	0	0
Hypothesis: non-catastrophic impacts equal Nordhaus and Sztorc (2013)'s estimate of 0.002132									
p-value	-	-	0.6364	0.0007	0.0009	0.0029	0.4023	0.3082	0.0752
Hypothesis: total impacts equal Nordhaus and Sztorc (2013)'s estimate of 0.002132									
p-value	-	-	-	-	-	-	0.0000	0.0000	0.0062
2015 SCC (2015 USD per metric ton of CO2e)									
Non-cat (5%)	-	$8	$12	$49	$47	$38	$14	$10	-$4
Non-cat (50%)	$22	$22	$25	$76	$74	$63	$28	$34	$95
Non-cat (95%)	-	$38	$39	$106	$104	$90	$43	$60	$245
Total (50%)	-	-	-	-	-	-	$68	$70	$140

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**See Tables 4 and 5 for notation.*

Table 7. Alternative Regression Applying More Restrictive Definition of Duplication: WLS with Cluster Robust Standard Errors at the Model Level, by dataset

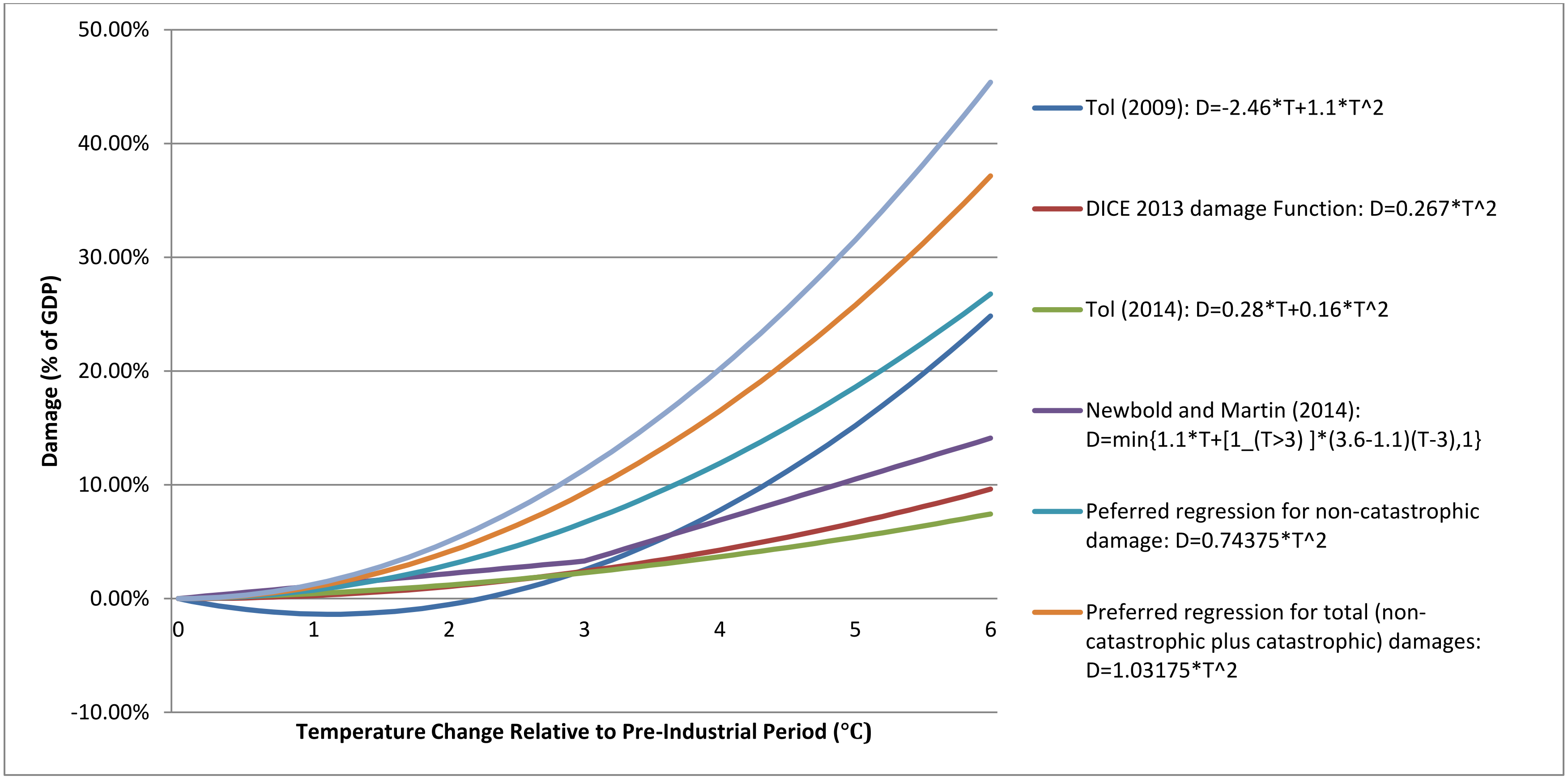
Dataset	Damages for Temp. Increases<3.2 °C		Damages for All Temp. Increases	
	Non-cited	Unique	Non-cited	Unique
	(2)	(4)	(6)	(8)
VARIABLES	D_new	D_new	D_new	D_new
t2	0.595**	0.601**	0.318**	0.601**
	(0.191)	(0.186)	(0.102)	(0.180)
mkt_t2	-0.680**	-0.727*	-0.403**	-0.727**
	(0.244)	(0.306)	(0.179)	(0.296)
cat_t2	0.288	0.390	0.364***	0.0202
	(0.293)	(0.289)	(0.103)	(0.185)
prod_t2	0.183	0.224	0.471	0.685*
	(0.153)	(0.298)	(0.271)	(0.358)
cross	3.162***	3.251**	3.162***	3.251***
	(0.915)	(0.928)	(0.883)	(0.896)
Observations	21	18	26	20
R2	0.759	0.773	0.873	0.911
Adjusted R-squared	0.683	0.686	0.843	0.881
Liklihood	-43.30	-37.51	-71.96	-50.39
F-statistic	15.61	13.81	572.3	219.0
Prob>F	0.000601	0.00306	0	2.47e-08
Hypothesis: non-catastrophic impacts (captured t2) equal Nordhaus and Sztorc (2013)'s estimate of 0.2132				
p-value	0.0803	0.0827	0.3291	0.0635
Hypothesis: total impacts (captured by t2 + cat_t2) equal Nordhaus and Sztorc (2013)'s estimate of 0.2132				
p-value	0.0072	0.0072	0.0000	0.0000
Hypothesis: non-catastrophic impacts (captured t2 + prod_t2) equal Nordhaus and Sztorc (2013)'s estimate of 0.2132				
p-value	0.0499	0.0920	0.0752	0.0183
Hypothesis: total impacts (captured by t2 + prod_t2 + cat_t2) equal Nordhaus and Sztorc (2013)'s estimate of 0.2132				
p-value	0.0078	0.0193	0.0062	0.0158

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Non-cited datasets define duplication as multiple estimates from the same model or estimates significantly based on already included estimates. In addition to the non-cited requirements, the unique dataset requires that no estimate is drawn from the same paper even if estimates apply to different temperature levels or apply different estimation methods*

Figures

Figure 1. Temperature-Damage Relationship for Previous Meta-Analyses and the Preferred Regression (Regression (4) on Table 4) from Our Study



*Following Nordhaus and Sztorc (2013), we multiple non-catastrophic and productivity impacts by 25% to account for potential omitted non-catastrophic impacts of climate change.

Figure 2. Box Plots of the damage estimates for the (a) "Final" and (b) "Non-cited" datasets with respect to (i) author and method, and (ii) model

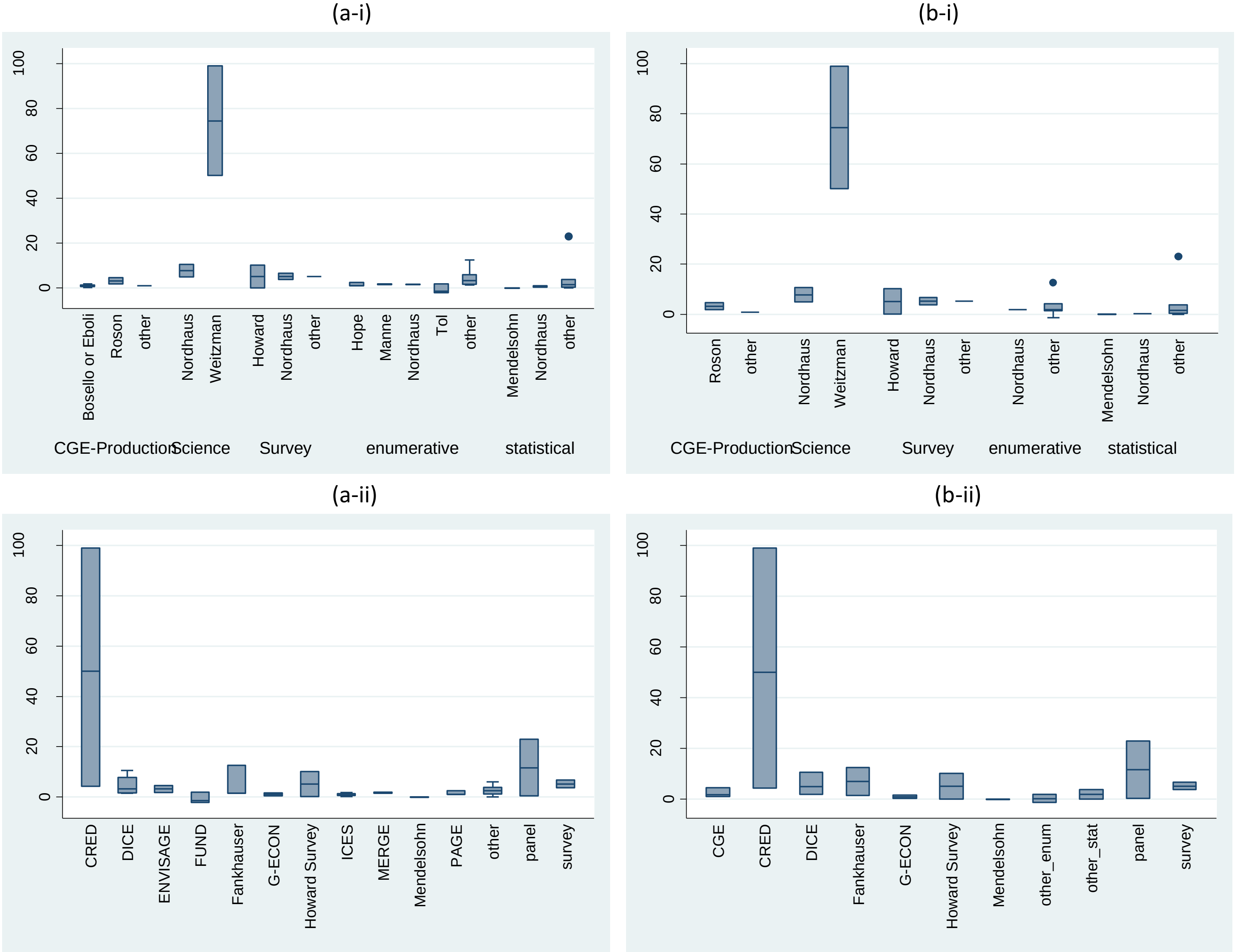


Figure 3. Scatter Plots of the damage estimates with respect to temperature change and time for: (a) “Final” dataset, and (b) “Non-cited”

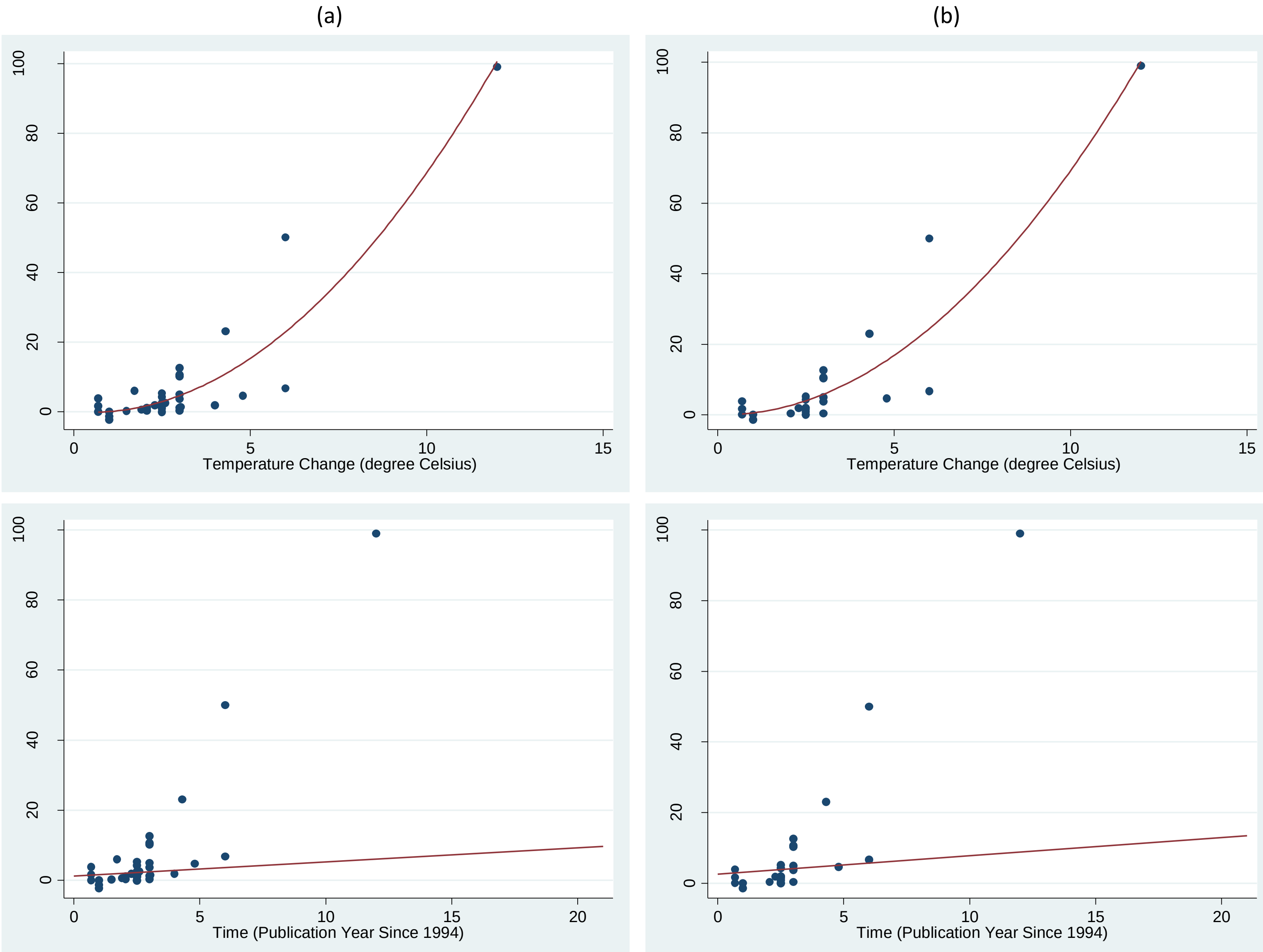


Figure 4a. Social Cost of Carbon over Time Calculated using Nordhaus and Sztorc (2013) and Our Preferred Regression (Regression (4) on Table 4), excluding climate impacts on economic productivity

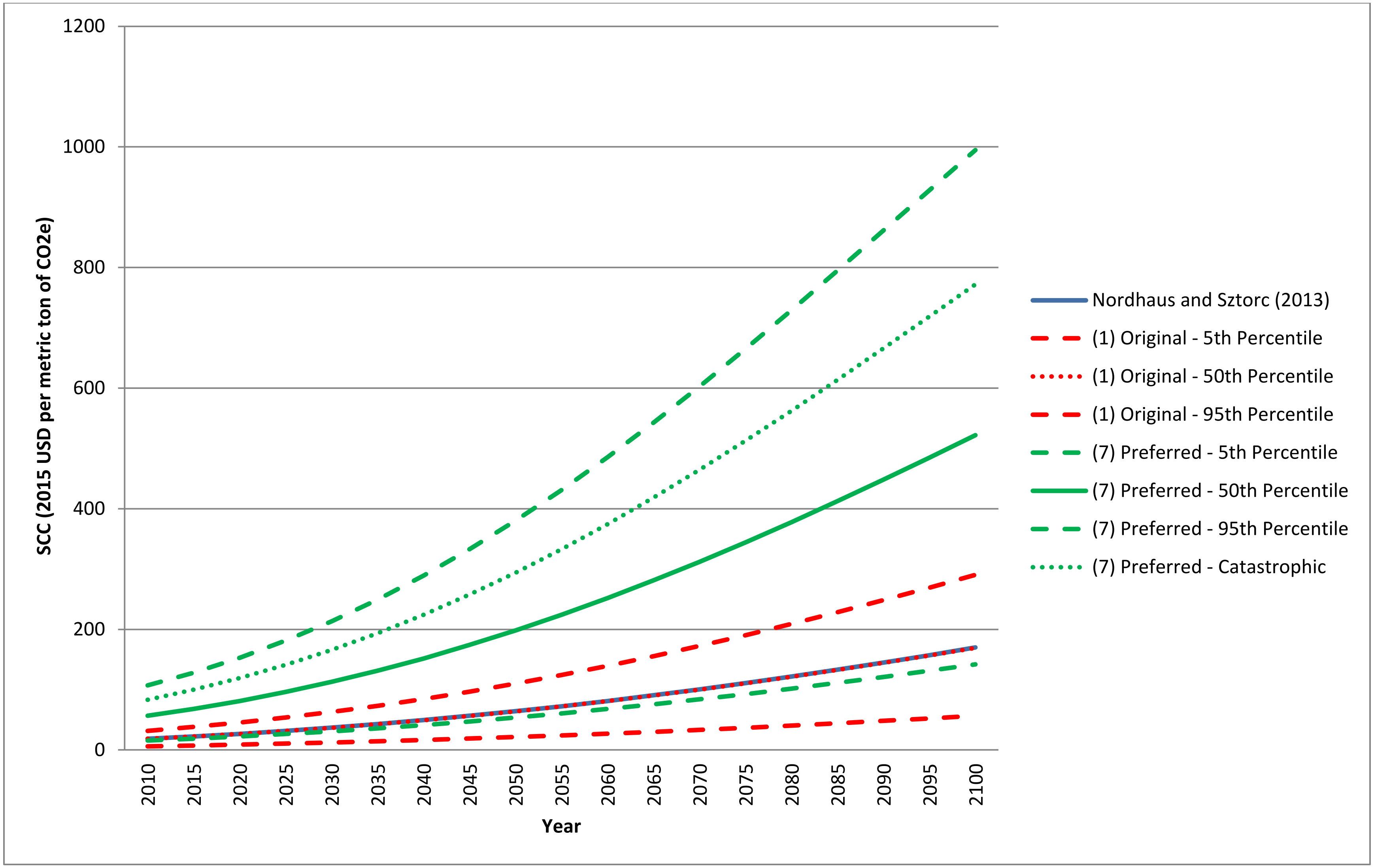
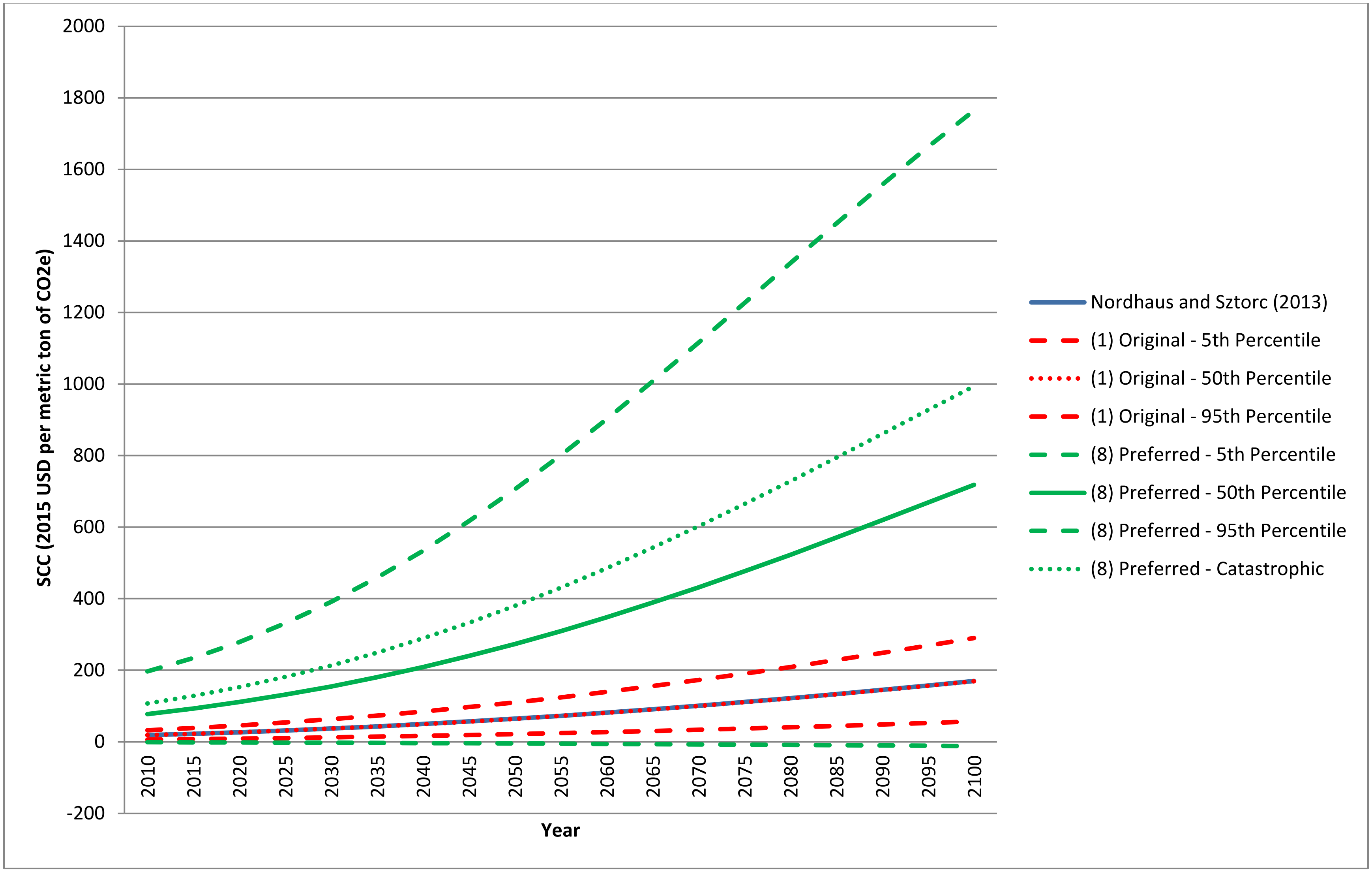


Figure 4b. Social Cost of Carbon over Time Calculated using Nordhaus and Sztorc (2013) and Our Preferred Regression (Regression (4) on Table 4), including climate impacts on economic productivity in non-catastrophic climate impacts



Endnotes

---

[1] The normality assumption may not hold – causing inefficient coefficient estimates and biased standard errors – if errors terms are right-skewed due to more negative than positive climate surprises (Tol, 2009).

[2] When measurement error is unobserved, some meta-analyses use sample size of study (n) to construct a weight $(\sqrt{n})$ because studies with larger sample sizes tend to be more accurate (Day, 1999; Nelson and Kennedy, 2009). In the case of climate change, the number of observations is unobserved for non-statistical based estimates.

[3] The search terms were: "percentage of GDP", "climate change", & global; "percent of GDP", "climate change", & global; "percent GDP", "climate change", & global; "% of GDP", "climate change", & global; "% GDP", "climate change", & global; "climate change" & "world output"; "Estimated impact of global warming on world output"; "climate change", "economic impact", & global; "climate change" & "global impact".

[4] When it was necessary to select only one estimate, we chose the estimate that utilizes the BAU scenario, the most climate information available (e.g. temperature and precipitation changes), and GDP weights.

[5] While testing for publication bias would be ideal, it is not possible in this context because many estimates do not provide standard errors. Thus, the Egger and Begg test statistics for publication bias cannot be constructed.

[6] As a consequence of dropping the compensating surplus estimates, there is reason to believe that the temperature-damage relationship captured by this study represents a lower bound estimate of the effects of climate on wellbeing. We analyze willingness to pay to avoid climate damages. However, willingness to accept estimates – which will imply higher damages per temperature increase – are more appropriate for regions that suffer economic losses because the current climate is "owned" by the current generation and they are being asked to accept a future climate that is less desirable. As a consequence, the handful of compensating surplus estimates – as estimated in the Maddison and Rehdanz studies – are more appropriate welfare measurements.

[7] Nordhaus (1994a) arbitrarily assumes that total U.S. climate damages will equal 1% of GDP for a 3 °C increase in global average surface temperature. Nordhaus and Yang (1996) is based on Nordhaus (1994a).

[8] We assume that (1) all scientific-based damage estimates account for catastrophic impacts, and (2) panel estimates of the temperature-damage relationship (Dell et al., 2012; Burke et al., 2015) capture the effects of climate change on GDP via productivity because they measure the effect of climate change on economic growth.

[9] Following Tol (2009), we regress damages on *Time* finding that damage estimates (insignificantly) increase over time. To check our results for robustness, we regress damages on time and time squared with and without controlling for temperature. In general, the time variables are insignificant except when we include time (linear term only) and temperature where we find that damage estimates significantly increase in magnitude over time.

[10] Assuming no initial benefits from climate change also preserves degrees of freedom.

[11] In our dataset, this adjustment is captured via *Alt_Curr_NASA* – which is a variable measured using a five year average of global annual mean land-ocean temperature index relative to 1880 (http://data.giss.nasa.gov/gistemp/graphs_v3/Fig.A2.txt).

[12] Given that many of these high estimates are also potential extrapolations of low temperature damages (i.e., all but one is drawn from a study with another included estimate), dropping these estimates also ensures greater independence and less duplication in the *low* dataset.

[13] The indicator variable for productivity is perfectly correlated with CGE and panel models, though not statistical models in general.

[14] We also find mixed evidence of non-normal error terms.

[15] We can utilize seemingly unrelated regression to test for statistical difference in the climate damage impacts when we account for outliers. In terms of statistical significance, only the coefficients corresponding to temperature-squared and its interaction with the market indicator variables significantly differ between the *low* and *all* datasets; this is due to the large uncertainty underlying productivity and catastrophic impacts and the near identical impact of the cross-sectional indicator variables over the two datasets. While the coefficient corresponding to non-catastrophic impacts differs (i.e., the coefficient corresponding to temperature-squared), all coefficients are only jointly significantly different between the *low* and *all* datasets for the shorter set of extended variables at the 5% level (i.e., that exclude productivity and cross-sectional dummies) and we cannot reject the null of equality in total damages (non-catastrophic plus catastrophic) across the two datasets.

[16] Using seemingly unrelated regression to provide a more robust test (relative to the tests in Table 4) of the null hypothesis that the non-catastrophic and total damage estimates differ between our preferred dataset (i.e., low,

non-cited) and Nordhaus and Sztorc (2013), we again reject the null at the 5% significance level, regardless of whether productivity is jointly considered. For the preferred specification using all temperature data (i.e., temperature changes below and above 3.2 °C), we generally reject the null hypothesis that the non-catastrophic and total damage estimates are statistically equivalent.

[17] Using seemingly unrelated regression, we reject the null that the coefficients corresponding to temperature squared and its interactions are equal across the *non-cited* and *final* datasets – i.e., when we do and do not account for duplication – at the 1% significance level when considering only damage estimates for low temperature increases (i.e., excluding outliers). At the 5% significance level, the difference is also significant when we consider only the coefficient corresponding to non-catastrophic damages (i.e., temperature squared) and total damages (i.e., temperature squared and its interaction with the catastrophic indicator variable), regardless of whether we consider productivity jointly. When we include outliers (i.e., use the *all* dataset), we fail to reject the null of no duplication for all, non-catastrophic damage, and total damage variables.

[18] Including an indicator variable for grey literature and/or its interaction with temperature-squared in all specifications in Table 4, we fail to reject the null of no publication bias.

[19] Using seemingly unrelated regression, we reject the null that the coefficients corresponding to temperature squared and its interactions are equal across the *non-cited* and *final* datasets – i.e., when we do and do not account for duplication – at the 1% significance level when considering damage estimates with and without outliers (i.e., *low* and *all*). Again at the 1% significance level, these results also hold when we consider only the coefficient corresponding to non-catastrophic damages (i.e., temperature squared) and total damages (i.e., temperature squared and its interaction with the catastrophic indicator variable), regardless of whether we consider productivity jointly.

[20] We find that our results for our preferred regression (i.e., regression (4) in Table 4 and regression (1) in Table A3) are relatively robust to various specification changes including: (2) reclassifying science-based damage estimates as not containing catastrophic impacts, (3) dropping science-based damage estimates, (3) assuming that CGE models that capture the impacts of ecosystem services on the market account for non-market impacts, (4) redefining the cut off for low temperatures as a 4.5 °C increase (approximately the temperature increase predicted by 2100 for the BAU scenario (RCP8.5) in the last IPCC (2013) report on the physical science of climate change), and (5) and (6) the use of different data to estimate the temperature adjustment terms.

[21] For our preferred dataset (i.e., the *low* dataset), we find that the magnitude of the catastrophic impact of climate change (*cat_t2*) is sensitive to the inclusion of Meyer and Cooper (1995) and one of the estimates in Howard and Sylvan (2015), while non-catastrophic impact of climate change (*t2*) is sensitive to the inclusion of the latter point only (in the opposing direction). If we drop both points, non-catastrophic impacts are diminished by 1/3 and catastrophic impacts increase by 1/10, though our results are much less sensitive to Huber weights.

[22] Most of the underlying estimates systematically omit key climate impacts that could significantly increase climate damages, including socially-contingent climate impacts (migration, social and political conflict, and violence), ocean acidification, etc. (Howard, 2014; Revesz et al., 2014). Given that some of the climate damage estimates in this study are over two-decades old and the positive coefficient corresponding to Time, we should expect climate damage estimates to continue to increase over time as currently omitted climate impacts are captured in future damage estimates. Given this bias, an adjustment upwards of the damage-temperature relationship – such as the 25% adjustment employed by Nordhaus and Sztorc (2013) – may be justifiable when using meta-analysis results to calibrate an IAM damage function.

[23] For purposes of sensitivity analysis, we apply a more restrictive definition of duplication. Specifically, the *unique* dataset applies a more extensive definition of duplication by further restricting data to one estimate per study (even if they are independent according to the author) from the non-cited dataset (Nelson and Kennedy, 2009). This latter definition requires us to choose which damage estimates to drop when two estimates are provided by a study. For CGE, enumerative, and survey based studies, we drop the estimate corresponding to a higher temperature increase assuming that they rely on extrapolation. For scientific studies, we maintain the estimate that is based on scientific principles. These two definitions of duplication reduce our dataset to 31 and 20 observations, respectively.

[24] With the exclusion of these potential duplicate estimates, the impact of outliers on the coefficient corresponding to temperature-squared is insignificant, though their impact on the coefficients corresponding to productivity and catastrophic impacts is potentially even greater.

[25] Another alternative, laid out by Kopp, Hsiang, and Oppenheimer (2013) is to develop an infrastructure that uses statistical (for example, Bayesian) methods to update damage functions as new estimates become available.