Title
Three Essays on Environmental and Development Economics

Author
Chong, Howard G.

Publication Date
2011

Peer reviewed|Thesis/dissertation

Three Essays on Environmental and Development Economics

by

Howard G Chong

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Agricultural and Resource Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Maximillian Auffhammer, Chair
Professor Severin Borenstein
Professor David Zilberman

Spring 2011

Three Essays on Environmental and Development Economics

Abstract

Three Essays on Environmental and Development Economics

by

Howard G Chong

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Professor Maximillian Auffhammer, Chair

This dissertation emcompasses three empirical studies in environmental and development economics. In Chapter 1, I study whether electricity use in newer or older residential buildings rises more in response to high temperature in a region of Southern California. Peak electricity demand occurs at the highest temperatures which are predicted to increase due to climate change. Understanding how newer buildings differ from older buildings improves forecasts of how peak electricity use will grow over time. Newer buildings are subject to stricter building energy codes, but are larger and more likely to have air conditioning; hence, the cumulative effect is ambiguous. This paper combines four large datasets of building and household characteristics, weather data, and utility data to estimate the electricity-temperature response of different building vintages. Estimation results show that new buildings (1970-2000) have a statistically significantly higher temperature response (*i.e.*, use more electricity) than old buildings (pre-1970). Auxiliary regressions with controls for number of bedrooms, income, square footage, central air conditioning, ownership, and type of residential structure partially decompose the effect. Though California has had extensive energy efficiency building standards that by themselves would lower temperature response for new buildings, the cumulative effect of new buildings is an increase in temperature response. As new buildings are added, aggregate temperature response is predicted to *increase*.

In Chapter 2, my co-authors and I investigate the effect of cap-and-trade regulation of CO_2 on firm profits by performing an event study of a CO_2 price crash in the EU market. We examine returns for 90 stocks from carbon intensive industries and 600 stocks in the broad EUROSTOXX index. Firms in carbon intensive, or electricity intensive industries, but not involved in international trade were most hurt by the event. This implies investors were focused on product price impacts, rather than compliance costs. We find evidence that firms' net allowance positions also strongly influenced the share price response to the decline in allowance prices.

In Chapter 3, my co-authors and I measure and examine data error in health, education and income statistics used to construct the Human Development Index. We identify three sources of data error which are due to (i) data updating, (ii) formula revisions and (iii) thresholds to classify a country's development status. We propose a simple statistical framework to calculate country specific measures of data uncertainty and investigate how data error biases rank assignments. We find that up to 34% of countries are misclassified and, by replicating prior studies, we show that key estimated parameters vary by up to 100% due to data error.

To my parents and sister

Ten Chong, Lily Chong, and Sylvia Chong

Contents

Acknowledgments

I am deeply humbled by the intellect and comraderie by which I am surrounded and to which I will always aspire.

Max Auffhammer has given tremendous advice and has helped me understand not only how to do good research, but also how to be a good academic. Not only have I learned how to make sure my results are correct and not overstated, but also how to communicate them effectively to the academic and general community. Max also advised me to apply for and join the UC Energy Insitute as a research assistant, a critical step in my intellectual development.

Severin Borenstein shepherded me to developing my interests and academic voice. Of the many lessons I have learned, perhaps the most important is that rigor is not a substitute for practical knowledge. Institutional details are important if you want to understand fully the underlying incentives and motivations of firms and consumers. Otherwise, GIGO.

David Zilberman, one of my earliest advisors, has been a great role model. He has given me and numerous other students sharp feedback on our research ideas, often citing his own prodigious work. Leading by example, he has shown me that academic curiosity is the cornerstone of a successful and fruitful academic career. For practical matters of how to frame research questions and what the big picture is, his guidance has been critical.

For the last three years of my graduate career, the UC Energy Institute has been the home where I've found an incredible microcosm of intellectual discourse around the lunchtable. I sadly will only realize just how insanely great it is as I leave this happy nest.

If I were to comprehensively describe the contributions of all the people who have helped me in my development, I would spend a long time writing this acknowledgement and it would still be incomplete. Here I will simply list and thank some of these people:

> Everyone at UC Energy Institute (now Energy Institute at Haas), including Max Auffhammer, Carl Blumstein, Severin Borenstein, Lucas Davis, Amiee Gee, Meredith Fowlie, Catie Hausman, Chris Knittel, Erica Meyers, Karen Notsund, Carla Peterman, Jim Sallee, Catherine Wolfram, and especially Koichiro Ito; my co-authors: Maximillian Auffhammer, James Bushnell, Erin Mansur, David Sunding, and Hendrik Wolff; ARE's excellent staff, professors, and graduate students, in particular Peter Berck, Diana Lazo, Leslie Martin, Shanthi Nataraj, Leo Simon, Gail Vawter, and Sofia Berto Villas-Boas; Matt Kahn and Edward Arens; and finally the many friends whose companionship has supported me through these endeavors, especially Lisa Bauer, Karen Bell, Shawn Chiao, Calico Goodrich, Ying Lee, James McBryan, Roger Mong, Jane Simon, Biana Yelent, and my broader dance community.

To you all, I give my heartfelt thanks. Namaste.

Chapter 1

Building Vintage and Electricity Use: Old Homes Use Less Electricity In Hot Weather

1.1 Introduction

Understanding the relationship between electricity usage and temperature, *i.e.* temperature response, is important for climate change policy and long-range electricity infrastructure planning. Residential buildings are a substantial contributor to CO2 emissions. In the US, residential buildings account for 21% of 2008 CO2 emissions (Environmental Protection Agency 2010), with about 50% of residential energy going to space heating and air conditioning (Energy Information Administration 2009). Furthermore, temperature increases from CO2 emissions will affect electricity demand through increased cooling loads, *i.e.*, air conditioning use. Electric power plant construction and infrastructure decisions are strongly driven by peak electricity demand which in California occurs during periods of highest temperature.

If new buildings have higher temperature response[1], then the average temperature response will increase as new buildings are added. Peak demand per household will also increase. Policies to reduce greenhouse gas emissions or reduce energy use often aim to decrease peak and total electricity demand.

Temperature response is better than total electricity use as a measure of the performance of buildings. As the component of electricity usage that varies with temperature, temperature response isolates factors such as the thermal performance of the building, the size of the building, and the thermostat preferences of occupants. In contrast, total electricity use

[1] In this paper, temperature response is defined as the percentage increase (relative to usage on a 65°F day) in electricity use due to a 1°F increase in temperature. Higher temperature response means more incremental electricity use.

conflates these factors with appliance ownership (e.g., more televisions) and other factors that don't depend on the building.[2]

Whether newer or older residential buildings in California have higher temperature response has not been studied using field data. California has had the most extensive energy efficiency standards in the United States applied to new buildings. Engineering models (e.g., Marshall and Gorin (2007); Abrishami, Bender, Lewis, Movassagh, Puglia, Sharp, Sullivan, Tian, Valencia and Videvar (2005)), predict strong reductions in energy use (both peak and total use) due to these standards, *ceteris paribus*, but other factors can offset these increases. The sign of the cumulative effect, measured as the difference between new and old buildings, is ambiguous. I use field data to estimate the temperature response across houses of different vintages.

This paper uses (household, monthly) field panel data on electricity use linked to building vintage and other building and household characteristics. Household electricity usage (quantity) data in Riverside County, California, USA, is regressed on time series variation in temperature to estimate temperature response. Cross sectional variation in building vintage and other characteristics at the Zip9-level or census block group-level identifies the temperature response by vintage.

The main finding is that each successive decade since 1970 has statistically significantly *increased* temperature response compared to older buildings (built prior to 1970). Hence, average peak load is expected to increase due to population growth and ensuing new construction. This exacerbates the impact of climate change on electricity use. Auxiliary regressions add controls for bedrooms, income, sqft, central air conditioning ownership, and type of residential structure. These differ across vintage and partially explain the increase in temperature response for newer buildings. With these controls, 1990s homes are estimated to have a temperature response of 8% less to 6% more than pre1970s homes in the most unrestrictive specification.

The organization of the paper is as follows. Section 2 presents existing related studies. Section 3 presents a description of the data. Section 4 presents an econometric model. Section 5 estimates the model. Section 6 discusses results and potential mechanisms. Section 7 concludes.

[2]Though I focus on temperature response, I also present comparisons of the total electricity use across vintage in Appendix 4.1.1. Unsurprisingly, new homes use more electricity, principally because they are larger.

1.2 Related Work

1.2.1 Temperature Response and Building Vintage in Field Evidence and Forecasting

Several papers have focused on temperature response of buildings using field evidence but have ignored how buildings have changed across vintage. Aroonruengsawat and Auffhammer (2009) examined the variation in the non-linear relationship between temperature and electricity use by sixteen climate zones in California, showing that the strongest relationships are in hotter inland areas. Earlier work on temperature response with (annual, state)-level data by Deschênes and Greenstone (2008) predicted that climate change scenarios generate a 33% increase in residential energy consumption nationwide *with the current set of buildings*. New buildings, if they perform worse than older buildings, may exacerbate this predicted increase.

By ignoring vintage effects, such studies would underestimate the impact of new buildings. Baxter and Calandri (1992) use an engineering model to estimate the impact of a 1.9°C temperature increase, finding a 2-4% increase in electricity use, but the study holds the building stock fixed. More recent work suggests that newer buildings are more temperature responsive. Every two years, the California Energy Commission runs a detailed simulation model to construct its demand forecast that includes a large mix of econometrically estimated parameters and engineering estimates. In a recent revision, they find that air conditioning saturation for newer buildings increased unexpectedly for both hotter (inland) and cooler (coastal) areas (Marshall and Gorin 2007). [3]

A limitation of engineering studies is uncertainty about whether engineering parameters represent actual field performance. Joskow and Marron (1992) describe many factors that contribute to overstatement of program effectiveness. In particular, a rebound effect may exist where occupants demand more services by responding to a decrease in the price due to efficiency (Greening, Greene and Difiglio 2000), interventions may imperfectly translate to the field, or unexpected confounding effects could diminish or accentuate savings. Although only a small portion of their broader critique, Joskow and Marron (1992) highlight the difficulty of extrapolating from the laboratory to the field. In Joskow and Marron (1993), they find that the ratio of measured to estimated savings are 0.31-0.42 for two 1980s retrofit programs; that is, engineering predictions overstated savings by a factor of 2 to 3. As more current evidence that field measurements and engineering estimates differ, Larsen and Nesbakken (2004) compare an econometric decomposition approach to the predictions of engineering models in Norway. They find that the two approaches decompose end uses quite differently. Hirst (1990) surveys the broader question of program evaluation. Nadel and

[3] Their large simulation model does not directly report temperature response. Instead, they report a related statistic, load factor, which is defined as average demand relative to peak demand. Load factor and average temperature response are inversely related. They project that load factor will decrease suggesting that newer homes should have higher temperature response.

Keating (1991) summarize results of a large number of field evaluations and find generally positive, but usually smaller, savings than what engineers predict. Use of field data, like that done in this paper, can produce more realistic forecasts or provide ways to validate engineering estimates. If engineering parameters overstate energy savings, then demand forecasts will be biased downward.

Two very recent papers use field data to test the impact of building vintage, both using monthly utility data. Jacobsen and Kotchen (2009) analyze one building standard code change in Florida using a sharp regression discontinuity. They estimate a 4-6% reduction is energy use. Costa and Kahn (2010) estimate the differences in total electricity use by building vintage for buildings in a community in California using cross-sectional variation and show that homes built after 1983 had lower total electricity use. My research looks at the differences for homes over three decades and focuses on differences in temperature response.

1.2.2 The Rosenfeld Curve and Energy Efficiency

Per capita total electricity sales for California have been relatively flat since the mid-1970s, when landmark legislation for energy efficiency was passed. Comparatively, sales for the rest of the United States have gone up by 50% (Figure 1.1). Explanations of this time series phenomenon, commonly referred to as the Rosenfeld Curve, vary widely. One obvious potential explanation points to California's policies, especially the establishment of building and appliance standards unique to California, which also began in the mid 1970s. However, correlation is not causation. The visual remarkableness of this curve is tempered when looking at comparable curves for nearby states. A look at analogous "Rosenfeld Curves" of *residential* electricity per capita over time for eight Western States (Figure 1.2) presents a quick visual contrast to California's impressive performance relative to the United States (Figure 1.1). Three other states (NV, OR, and WA) have had flat residential electricity per capita profiles, though they had weaker building standards.[4]

Avoiding many of the problems of state-level analyses, my research uses rarely available microdata at the household-level with covariates at the 5-10 household-level. State-level analyses are problematic because they assume comparability across states. The identifying assumption in such studies is that changes in per-capita electricity load across states would have been the same in the absence of energy efficiency policies. This assumption is embedded in several state-level analyses: Aroonruengsawat, Auffhammer and Sanstad (2009) and Horowitz (2007) use state-level panel data; Sudarshan and Sweeney (2008) make a comparison between the US and California; and Loughran and Kulick (2004) and Auffhammer,

[4]Historical information for all states on building energy standards comes from the Building Codes Assistance Project (n.d.). Nevada implemented a mandatory building energy code in 1978 but "between 1983 and 1986, the state did not support or enforce this energy code". Oregon implemented a building energy code in 1978 that did not apply to residential buildings. A residential code was adopted in 2003. Washington adopted a voluntary energy code in 1977, with a mandatory code established in 1986.

Blumstein and Fowlie (2008) use utility-level panel data. These analyses typically find evidence that energy efficiency programs reduce energy consumption. However, the underlying assumption of comparability across states can be violated for many reasons. The evolution of a state's aggregate energy efficiency (as measured by residential electricity per capita) may depend on changes in the composition of the type of housing (urban vs rural, single family vs multifamily/mobile homes), differential growth in the size of housing, changes in geographic/climatic composition (e.g. coastal vs. inland), and differences in the adoption of air conditioning.

This analysis makes an important contribution to studies of policies aimed at reducing residential energy. In the context of "energy intensity" measures, such as electricity per capita or per GDP, my research identifies the *new and counterintuitive* empirical fact that households in new buildings use more electricity per household, both in total use and in response to temperature. It runs counter to what one might expect from looking at the Rosenfeld Curve, where per capita electricity has been flat, but the Rosenfeld Curve is an aggregate-level result that may conflate other factors.[5] Explaining what causes this empirical fact is important for understanding the effectiveness of building energy use policy in the context of many *simultaneous* changes.

1.3 Description of the Data

Three investor-owned utilities (Pacific Gas and Electric, Southern California Edison, and San Diego Gas and Electric) gave researchers at the University of California Energy Institute the complete billing history for all residential household bills in these electricity service territories. Time coverage for utilities varies, but the longest period of data are from 1998 to part of 2009 for Southern California Edison (SCE). Information includes billing start date, billing end date, total electricity used (kWh), total bill, an anonymized account id, an anonymized physical location id, and the zip code (usually at the nine digit level). This paper currently focuses on one county, Riverside County, where there are over 20 million observations for SCE customers.

Riverside County was chosen because it is an inland area with a wide range of temperatures, there is considerable variation in the building vintage, Aroonruengsawat and Auffhammer (2009) found this region to have substantial average temperature response, and detailed county assessor's property information is available. It is important to restrict to one county or area because housing design, climate, and building standards differ strongly across the state. For cleaning, bills with 25 days or less or 35 days or more were dropped (about 5%). Bills with less than 2kWh/day or more than 80kWh/day are outliers were also

[5]Given that households in new buildings use more electricity than those in older buildings, if older buildings have not changed, it follows logically that the average household use would go up. Since this contradicts the flat average electricity use (Rosenfeld Curve), the inference is that households in old buildings use significantly less electricity, to the point that the average use is flat.

dropped (about 4%).[6]

The billing data lacks housing and household information; two data sets of different spatial resolution are used to provide this information. County assessor's data (County of Riverside Assessor's Office 2010) was obtained for single family homes identifiable to the address. Because SCE billing includes both bills for single family and multifamily (e.g., apartments), I condition on census block groups where more than 95% of households are in single family homes. Bills are next matched to assessor data via the zip9. Zip9s are very small, with an average of 4.8 assessor records per zip9. For each zip9, the proportion constructed in each vintage category, the median of square footage, and the proportion of houses with central air conditioning for each zip9 is associated with all the bills in that zip9.

The second source of housing information is the US Census. The 2000 US Census's Summary File 3[7] (United States Census Bureau 2009) has at the census block group-level proportions of the vintage of housing, proportions of type of structure (single family vs multifamily vs mobile home), the number of rooms, and the income distribution. A census block group has a size on the order of 500 housing units. Figure 1.3 has a map of part of Riverside County by census tract.[8]. The shading corresponds to the proportion of housing in a tract that was built after 1980, with darker meaning more new construction. Hence, within this county, there is substantial spatial variation in the age of housing which is needed for estimating vintage differentiated temperature response; *i.e.* temperature response is compared between dark and light areas of the map. Because of the large number of observations and computing limitations, a 1-in-5 subsample was used to reduce the sample to 5.3 million observations when using census data.

Daily maximum (Tmax) and minimum (Tmin) temperature at a 4km x 4km grid are generated according to the algorithm used by Schlenker and Roberts (2009) which has been used for estimating the relationship between crop yields and temperature. The reader is directed there for a more full description of the algorithm as well as diagnostics that show the methodology is reliable. Billing data are then matched via Zip9 to the gridded temperature data and to the census block group. The average of Tmax and Tmin is then taken as the daily temperature. These are then translated into cooling degree days (CDD) and heating degree days (HDD) with a reference temperature of 65°F. In a more flexible approach which follows Aroonruengsawat and Auffhammer (2009), the daily temperature is binned into 10 bins with approximately equal number of observations. Temperature bin ranges are listed in Table 1.1.

To give a better sense of the data, Figure 1.4 gives plots of average daily electricity use versus time from the monthly billing data for one household. Peaks for electricity use correspond to summer months. This data is then replotted as average daily electricity use versus average cooling degree days in Figure 1.5. As temperature increases, the electricity

[6]The main analysis was rerun with a cutoff of 200kWh/day rather than 80kWh/day. Results did not markedly change.

[7]This is also known as the Census Long Form.

[8]On average, a census tract is 3 census block groups.

use for this household increases.

Summary statistics of the data (using assessor's data which is restricted to single family homes at the zip9 level) are in Table 1.2. Most homes (88%) have central air conditioning; the newest homes almost always have central air conditioning, but less than half of older homes have central air conditioning.

Summary statistics of the data (using census block groups) are in Table 1.3. The top section reports information from the billing data. The average household use per day is 25.5kWh, or 9307kWh per year. This is slightly lower than the national average of 11,500 kWh per year (Energy Information Administration 2009). The second section of the summary statistics corresponds to building and household characteristics from the Census data at the level of the census block group. 20% of observations were built in 1970-1979, 36% in 1980-1989, and 21% in 1990-2000, and 23% before and including 1969. The min and max of these variables are close to zero and one, which means there is substantial variation across census block groups in building vintage. The vintage variables differ from the previous table because this data set includes non-single family homes. The average number of bedrooms and rooms are 2.57 and 5.23, and the average household income is $48,200.

An extended data discussion with additional detail on data cleaning and matching is in Appendix 4.1.4.

1.4 Econometric Model

The average temperature response for subareas of California has been estimated by Aroonruengsawat and Auffhammer (2009) and nationally by Deschênes and Greenstone (2008). A similar estimating equation is given by Equation 1.1. This flexibly estimates the average temperature response in log terms within the sample area after controlling for a household fixed effect.[9] Temperature is binned. D_{pit} is a scalar $[0, 1]$ that denotes the fraction of days where a household is exposed to the pth temperature bin.

$$\begin{aligned} ln(kWh\_useperday_{it}) = \;& \textstyle\sum_{p=1}^{BINS} \rho_p * D_{pit} \\ & +\alpha_i + \varepsilon_{it} \end{aligned} \tag{1.1}$$

An alternative specification is to parameterize the temperature response in terms of cooling degree days and heating degree days[10]. Following Reiss and White (2008), I include linear and squared terms for CDD and HDD which results in Equation 1.2.

[9]Studies relating energy use and temperature have varied in the functional forms used. I discuss this in Appendix 4.1.2. In the robustness checks and the auxiliary regressions, I include alternative functional forms.

[10]Degree days are referenced to 65°F. For a given day, $CDD = max(Tmean - 65, 0)$ and $HDD = max(65 - Tmean, 0)$

$$\begin{aligned} ln(kWh\_useperday_{it}) &= f(CDD, HDD) + \alpha_i + \varepsilon_{it} \\ &= \beta_1 CDD_{it} + \beta_2 CDD_{it}^2 + \beta_3 HDD_{it} + \beta_4 HDD_{it}^2 \\ &\quad + \alpha_i + \varepsilon_{it} \end{aligned} \tag{1.2}$$

I have estimated both temperature parameterizations. The degree day parameterization is the main specification presented. A limited number of binned results are also presented.

I next estimate the heterogeneity of temperature response by vintage. The vintage of each household is not known, but the proportion of buildings of each vintage in an area is known, either at the Zip9- or census block group-level. The temperature response of each vintage is estimated via the cross sectional variation in vintage across areas. Equation 1.3 uses the degree day parameterization, while Equation 1.4 estimates the average response by vintage using binning.

$$\begin{aligned} ln(kWh\_useperday_{ijt}) &= \sum_{v=1}^{VINTAGES} V_{jv} * (\beta_{1v} CDD_{it} + \beta_{2v} CDD_{it}^2 + \beta_{3v} HDD_{it} \\ &\quad + \beta_{4v} HDD_{it}^2) + \alpha_i + \varepsilon_{it} \end{aligned} \tag{1.3}$$

$$ln(kWh\_useperday_{ijt}) = \sum_{v=1}^{VINTAGES} \left(\sum_{p=1}^{BINS} [\beta_{pv} V_{jv}] \right) * D_{pit} + \alpha_i + \varepsilon_{it} \tag{1.4}$$

where

- i, j, t index households, Zip9 or census block groups, and time (monthly billing period), respectively
- $BINS$ represents the number of temperature bins, p indexes them.
- $VINTAGES$ represents the number of building vintage categories. v indexes them.
- V_{jv} is in [0,1] and represents the proportion of buildings in j for vintage v
- D_{pit} is in [0,1] and is the measure of the proportion of days for household i in the billing cycle t where the average temperature is in the pth bin

In both regressions, the mean temperature-invariant consumption is captured by the household fixed effect, α_i. Importantly, this will flexibly capture temperature invariant factors such as variation in appliance ownership and usage patterns.[11] In Equation 1.4, the

[11] A more common specification would also include time dummies. This specification has been run with both month(Jan-Dec) and year dummies. The pattern of results is the same for the CDD and HDD parameterization.

parameters of interest are the β_{pv} that represent the temperature response for the pth temperature bin for the vth vintage.[12] The set of β_{pv} plotted against the p temperature bins yields the temperature response. Electricity use should increase with increasing temperature, represented by $\beta_{p^*v} > \beta_{p'v}$ when p^* is hotter than p' in the air conditioning range of temperatures for a given v.[13] If new buildings have higher temperature response than older buildings, then $\beta_{pv^*} > \beta_{pv'}$ when v^* is newer than v' for any p in the air conditioning range of temperatures. In Equation 1.3 with the degree day parametrization, the β_{1v} and β_{2v} determine the temperature response to hotter temperatures. Temperature response is higher when these coefficients are larger. In the degree day parameterization, the comparison of interest is the analogous differences in predicted temperature response across vintages.

Estimation of Equations 1.3 and 1.4 determines the average temperature response by vintage but does not identify the causal effect of building standards. Over time, buildings have changed in numerous ways, such as building standards on insulation and glazing, efficiency standards on appliances, the likelihood to have air conditioning, the square footage, and building design. The standard practice of using $ln(kWh\_useperday)$ as the dependent variable is one way to controls for square footage and size as discussed in Appendix 4.1.2, but the other factors are captured by the vintage effect. Building standards do vary by vintage and are predicted via engineering estimates to have an impact on temperature response. However, building standards cannot be isolated from the other changes.[14] Hence, I interpret the estimate to Equations 1.3 and 1.4 as the cumulative impact of multiple changes.

In order to aid interpretation of the cumulative effect, available covariates can be added which can isolate some factors of the cumulative impact of vintage, but the remaining factors cannot be isolated. County assessor's data provide additional covariates for central air conditioning ownership and square footage at the zip9-level but only for areas almost entirely composed of single family homes. Using this data, the following auxiliary specifications can be estimated, the first with the degree day parameterization and the second with temperature bins. Importantly, building standards are not controlled for and would still be part of the vintage effect.

[12] One of the temperature bins, $62.7°\text{F} - 66.4°\text{F}$ is left out wlog as the reference temperature bin, otherwise the rank condition is violated.

[13] The heating range of temperatures is estimated but not discussed in this paper. Heating fuel varies across vintage, with newer homes more likely to have natural gas as their primary heating fuel. In contrast, electricity is almost universally the energy source for cooling.

[14] There are two potential methods of estimating the causal impact of building standards. First, a regression discontinuity (RD) design may be possible if the treatment is discontinuous. However, building standards implementation could be slow and gradual, which would not be picked up by an RD design. Jacobsen and Kotchen (2009) apply an RD approach which assumes a sharp change in standards implementation. Second, cross state comparisons can be made, but the limitations of cross-state analyses has been discussed.

$$\begin{aligned} ln(kWh\_useperday_{izt}) \quad = \quad & \sum_{v=1}^{VINTAGES} V_{zv} * (\beta_{1v}CDD_{it} + \beta_{2v}CDD_{it}^2 \\ & \qquad +\beta_{3v}HDD_{it} + \beta_{4v}HDD_{it}^2) \\ & +CentralAC_z * (\varphi_1 CDD_{it} + \varphi_2 CDD_{it}^2 \\ & \qquad +\varphi_3 HDD_{it} + \varphi_4 HDD_{it}^2) \\ & +SquareFootage_z * (\theta_1 CDD_{it} + \theta_2 CDD_{it}^2 \\ & \qquad +\theta_3 HDD_{it} + \theta_4 HDD_{it}^2) \\ & +\alpha_i + \varepsilon_{it} \end{aligned} \quad (1.5)$$

$$\begin{aligned} ln(kWh\_useperday_{izt}) = \quad & \textstyle\sum_{p=1}^{BINS} \Big(\displaystyle\sum_{v=1}^{VINTAGES} [\beta_{pv}V_{zv}] + \\ & \varphi_p CentralAC_z + \\ & \theta_p SquareFootage_z \Big) * D_{pit} + \\ & \alpha_i + \varepsilon_{it} \end{aligned} \quad (1.6)$$

where

- i, z, t index households, zip9, and time (monthly billing period), respectively,
- V_{zv} is in [0,1] and represents the proportion of buildings in z for vintage v
- $CentralAC_z$ is the proportion of buildings with central air conditioning in z, and
- $SquareFootage_z$ is the median square footage for buildings in z.

With the census data, three variables are interacted with temperature response that vary at the census block group-level: (1) average ln(income), (2) average number of bedrooms (a proxy for size), and (3) the type of structure, *i.e.* Single Family or Multifamily or Mobile/Other. Equation 1.7 presents this auxiliary specification with the degree day parameterization.

$$
\begin{aligned}
ln(kWh\_useperday_{ijt}) \quad = \quad & \sum_{v=1}^{VINTAGES} V_{jv} * (\beta_{1v}CDD_{it} + \beta_{2v}CDD_{it}^2 \\
& \qquad\qquad + \beta_{3v}HDD_{it} + \beta_{4v}HDD_{it}^2) \\
& + \sum_{s=1}^{STRUCTURES} STR_{js} * (\rho_{1s}CDD_{it} + \rho_{2s}CDD_{it}^2 \\
& \qquad\qquad + \rho_{3s}HDD_{it} + \rho_{4s}HDD_{it}^2) \\
& + AvlnIncome_j * (\gamma_1 CDD_{it} + \gamma_2 CDD_{it}^2 \\
& \qquad\qquad + \gamma_3 HDD_{it} + \gamma_4 HDD_{it}^2) \\
& + AvBedrooms_j * (\delta_1 CDD_{it} + \delta_2 CDD_{it}^2 \\
& \qquad\qquad + \delta_3 HDD_{it} + \delta_4 HDD_{it}^2) \\
& + \alpha_i + \varepsilon_{it} \qquad\qquad (1.7)
\end{aligned}
$$

STR_{js} is in [0,1] and represents the proportion of buildings in j for the type of structure, s. $AvlnIncome_j$ is the average of ln(income) per household in j. $AvBedrooms_j$ is the average bedrooms per household in j. j indexes census block groups areas. Importantly, building standards and measures of air conditioning ownership are not available as covariates when using census data.

An important property of these estimates of temperature response is that they are immune to many types of omitted variable bias. In order for omitted variable bias to bias temperature response results, two conditions must be met. First, the omitted variable must vary across vintage. Second, the omitted variable must be correlated with temperature. A variable, such as price, that does not vary within this region nor by temperature, would not bias results, except if price elasticities for cooling varied across vintages.[15] Aroonruengsawat and Auffhammer (2009) included price as a regressor in estimating regional temperature response and found that it did not affect results.

1.5 Results

1.5.1 Main Results: Degree Day Parameterization With County Assessor's Data

Results presented in this subsection use the degree day parameterization and county asses-

[15] This is a future piece of planned work. It is not straightforward to include prices because they are nonlinear, increasing block rate tariffs.

sor's data. Alternative specifications follow this subsection.

I first estimate the average temperature response across all households given earlier by Equation 1.2. Column A1 of Table 1.4 and Figure 1.6 present the results of the estimation using fixed effects panel regression with standard errors clustered at the zip9 level. This shows the strong increase in electricity in response to temperature for higher temperatures, relative to 65°F.[16]

Next, I estimate temperature response *by vintage* as given earlier by Equation 1.3. Column A2 of Table 1.4 and Figures 1.7 to 1.10 present the results of the estimation. The omitted vintage variable is pre1970s, so the coefficients on the remaining variables are differences from the temperature response of pre1970s buildings. Figure 1.7 is the temperature response for pre1970s buildings. Figures 1.8 to 1.10 are for each other vintage relative to pre1970s buildings. Each figure has a horizontal line at zero to indicate what would result if there were no difference between vintages. To interpret these results, the 1970s, 1980s, and 1990s vintage of buildings have statistically significantly higher temperature response than pre1970s buildings. The highest temperature response is for 1990s buildings, followed by 1980s buildings, 1970s, and then pre1970s buildings.

Lastly, I estimate temperature response by vintage with some controls interacted with temperature response, as given by Equation 1.5. These controls capture variation in temperature response that is correlated with central air conditioning and square footage. Results are in Column A3 of Table 1.4 and Figure 1.11 which combines the graphs. Central air conditioning strongly positively increases temperature response and is more prevalent in newer buildings. Square feet negatively impacts CDD; this means that the percentage increase in electricity on a hot day is systematically *less* for larger buildings. This makes sense from an engineering perspective because a doubling of sqft typically would mean a less-than-doubling of surface area. As discussed in the Appendix 4.1.2, the main econometric specification assumes comparability across households of different size by comparing percent changes. In the figure, all of the temperature response curves shift downward because new buildings more often have air conditioning. 1970s buildings are not statistically significantly different from pre1970s buildings after adding controls. 1980s and 1990s buildings are still more temperature responsive after adding controls.

1.5.2 Robustness checks

To partially guard against the possibility that some of these results are driven by parametric assumptions on size, I re-estimate the previous regression and restrict square footage to 1300 to 1600 square feet which reduced the observations by about two-thirds. Estimation results

[16]The heating curves or temperatures below 65°Fare not included because electricity is not the dominant heating fuel. Hence estimated differences across vintages will be partly driven by differences in heating fuel. Reliable statistics of heating fuel across area are not available. Natural gas is more common in newer buildings and in some areas. Heating is still included to improve model fit and reduce standard errors.

are presented in Column A4 of Table 1.4 and Figure 1.12. The signs of the Square Foot $\times$ CDD and Square Foot $\times$ CDD^2 variables change, but it is also less statistically significant. Even with this change, cumulative responses by vintage with controls are similar to the main results.

The degree day parameterization may be overly restrictive. I run analogous regressions but with temperature binning instead of the degree day parameterization. Equation 1.4 presents the regression without controls. Results are given in Figure 1.13. Equation 1.6 presents the regression with controls. Results are given in Figure 1.8.[17] Results are similar to the main results. Without controls, all vintages have statistically higher temperature response for bins higher than 65°F. With controls, 1970s buildings are not statistically significantly different from pre1970s buildings for all bins, and 1980s and 1990s buildings are more temperature responsive.[18]

In each of these cases, the $ln(kWh_{perday})$ specification compares households in terms of the percent change in electricity use relative to each house's fixed effect, *i.e.* their temperature invariant mean usage. An alternative approach is to compare each household's temperature response in levels (as opposed to percentages) and control explicitly for size. This alternative is discussed and estimated in Appendix 4.1.2. Referring to Figure 1.15, this parameterization shows that the predicted temperature response for all vintages of buildings are not statistically significantly different from the reference group of pre-1970s buildings. Standard errors are larger due to the reduction in observations.

An alternative data source is census data which offers some advantages. Census data is not restricted to single family homes and includes income and other socioeconomic information. The disadvantage is that census block groups are larger geographically, so there is less spatial variation and more potential for bias from aggregation, as discussed in Appendix 4.1.3. Regressions are run with census block data. Figure 1.16 shows the temperature response by vintage after estimating Equation 1.3. 1980s and 1990s homes have a higher temperature response that is not statistically significantly different from pre1970s homes, but 1970s homes have a lower temperature response. Standard errors are much larger due to the decrease in number of areas. There are 372 census block group areas compared to 9316 Zip9's areas. Figure 1.17 shows the the results of estimating temperature response by vintage with controls for income, size, and type of structure, as described in Equation 1.7. Note that air conditioning is not available at this spatial resolution and is not used as a control. With controls, results change dramatically. 1970s buildings have a higher temperature response that is not statistically significant. 1980s and 1990s buildings have a higher temperature response that is statistically significant. The reason for the upward shift is that 1970s and 1980s buildings had a higher proportion of multifamily and mobile home units which have lower temperature response. After controlling for this, both curves shift upward. For the

[17] Regression tables available upon request.

[18] Note that caution should be used when looking at the lowest and highest temperature bins. These bins contain outliers and the intra-bin temperature distribution across vintages is quite large. Newer buildings have more data points in the highest temperature bin.

1990s buildings, income has a negative effect on temperature response and households in newer building have higher income. After controlling for this, the the 1990s curve shifts upward.

Total usage is another way to compare electricity use across households. This research focuses on temperature response under the argument that temperature response isolates elements of the building and household preferences only for cooling and heating services. In contrast, total usage captures many other differences across vintages, such as the number and type of appliance. Appendix 4.1.1 discusses this in more depth. The results, as presented in Table 1.5, show that new homes use statistically significantly more electricity than older homes in total electricity use (Column T1). This is expected since new homes are larger.

After adding controls for square footage and central air conditioning (Column T2), new houses use statistically significantly less electricity. Further adding controls for temperature across vintages (Column T3), new homes still use less electricity, and the coefficient on central air conditioning without temperature interactions becomes statistically insignificantly different from zero. This empirical result can be justified without invoking increased efficiency; a home with twice the square footage may not have twice the amount of people or appliance usage. Hence, the lower electricity per sqft is consistent with fewer services per square foot.

1.6 Policy Significance and Potential Mechanisms

1.6.1 Policy Significance

The results show that, in Riverside County, the cumulative temperature response for buildings has been stronger for newer buildings (1980s and 1990s) than for older buildings (1970s and pre1970s). This has two main policy impacts, one for load forecasting and one for the impacts of climate change given that the composition of the building stock is changing to something more temperature responsive.

First, in conducting load forecasts, these results suggest that new construction will increase the average temperature response and increase peak load on the hottest days. As a calibration, the population forecasts of RAND (RAND California 2010) predict an average annual population increases of 2.6% for Riverside County. Applying this growth to Riverside County and assuming that new construction has the same temperature response as 1990s buildings, Figure 1.18 predicts the increase in average temperature response on a 75°F day to go from 48.8% to 52.3% from today to 2020. Peak demand will increase proportionately as well. This is comparable to the estimated 3.7% increase in peak demand due to a 1.9°C increase in temperature as estimated by Baxter and Calandri (1992).

Looking at the issue of air conditioning statewide potentially could have an even greater effect. This is because coastal areas have historically had a lower amount of air conditioning, but the CEC revised forecast commented that there was an unexpected increased air conditioner saturation in cooler areas. Table 1.6 presents air conditioning saturation for old versus new housing by forecast climate zones from KEMA-XENERGY (2004) data. Figure 1.19 gives a map of the zones. Coastal areas that have very low ownership of air conditioners for older buildings have dramatically increased air conditioner ownership for newly built buildings.

Second, climate change impacts will be exacerbated with the increased temperature response from newer houses. Using the same calculation as given in Figure 1.18 above, I can predict the difference in climate change impacts adjusting for the estimate that new buildings are more temperature responsive. In 2050, Riverside's population is predicted to more than double. For a 5°F increase due to climate change, temperature response will be about 2-3% higher with the addition of new buildings compared to the current building stock.

1.6.2 Potential Mechanisms

As previously discussed, it is not possible to separate out the mechanism of the vintage-differentiated temperature response. The heterogeneity by vintage was first estimated, and then controls for observables were included, which captured some of the heterogeneity. The remaining temperature response is from the other factors.[19] One of the remaining factors that are part of the vintage temperature response coefficients were policy developments. This would include building standards implemented in 1975, 1979, 1984, and 1992 and appliance standards implemented in 1978 and 1987.

After controlling for differences in air conditioning, the remaining differences across households of different vintages is smaller and depends on the specification used. In the main log specification with controls for central air conditioning ownership, new buildings had statistically significantly higher temperature response by a small amount (Figure 1.11). Using a level specification and restricting the sample to houses of similar size, new homes performed slightly better, but not statistically significantly so (See Appendix 4.1.2, Figure 1.15).

Engineering estimates provide a prediction of the impact of buildings standards absent any other changes. Building standards have also varied by vintage and are predicted to reduce temperature response significantly by 34-56% for new versus old buildings. The CEC identifies four significant changes in building standards and estimates the savings from those standards with engineering models (Marshall and Gorin (2007) and Abrishami et al. (2005)). I summarize and report the savings from in Table 1.8. Total load reduction is about 6% from engineering estimates. However, to make this result comparable to my estimates, two adjustments must be considered. First, building standards only affect new construction and major renovation; these are represented in the fourth column which has the population increase since the standard went into effect as a proportion of the current population. Also, building standards only affect the temperature response component of electricity use. I

[19] This relies on the assumption that the other factors are uncorrelated. Otherwise, the included controls would pick up other factors through correlation with omitted variable.

calculate the implied reduction in temperature response, -34% to -56% from each building standard in the last column.

The juxtaposition of similar temperature response across vintages and a large predicted decrease in temperature response due to building standards suggests that other factors have had a large positive effect for new houses. There are multiple potential mechanisms, none of which the data can separate out. Behavioral responses, such as those driven by the rebound effect (Greening et al. 2000) can increase temperature response. This would mean that part of the increase is due to an increase in comfort from using more cooling services. New buildings may differ in their thermal design in that they may have taller ceilings, fewer trees, less passive shading, more structural complexity, or a higher window-to-wall ratio; all of which may increase the electricity needed to cool a building. It is also possible that there is sorting, where people who favor more cooling services are more likely to live in new buildings. Another possibility is that standards may not have been as effective as they have claimed, following the logic of Joskow and Marron (1992). These are factors that would need to be carefully considered when designing and evaluating of building standards.

Some auxiliary information suggests that sorting plays a limited role in explaining the results of higher temperature response in new buildings. Using data from KEMA-XENERGY (2004) for homes in this region, Table 1.9 shows the *self-reported* proportion of homes who turn on their air conditioning by vintage and time of day, and Table 1.10 the *self-reported* average thermostat set point conditional on having central air conditioning on by vintage and time of day. The newer buildings tend to turn on their air conditioner slightly more often, but the set point of the thermostat is not very different across vintages. This data cannot be used in the regression framework because it is available only for large areas whereas assessor and census data were available for small areas.

1.6.3 Future Work

Billing data are available for a large portion (about 80%) of California and future work will estimate this specification across the entire state. Though the average temperature response in coastal areas is low, according to Aroonruengsawat and Auffhammer (2009), the CEC reports suggest that new construction in lower temperature areas on the coast has had higher than anticipated air conditioning ownership. In fact, Table 1.6 shows that air conditioning ownership has increased strongly in both coastal and inland areas. Estimation of the entire state would enable me to aggregate county-level estimates to a statewide average cumulative temperature response.

This research also presents a puzzle about the causes of the Rosenfeld Curve, shown in Figure 1.1. Since the mid 1970s, per capita electricity consumption for California has been flat while it has increased 50% for the United States. The breakpoint in the 1970s coincided with the establishment of aggressive energy efficiency policies. The Rosenfeld Curve coupled with engineering estimates suggest that California's policies have been very effective, but this research suggests that, in terms of temperature response, the net effect has been that newer buildings increase temperature electricity use more than older ones in response to high temperatures in Riverside County, one of California's hottest counties. Several other drivers (most notably, population growth biased toward hotter areas which have higher electricity use) would also increase aggregate per capita electricity consumption. The resulting puzzle is why California has had a flat per capita electricity profile despite these drivers that would strongly push electricity use upwards. To try to understand the aggregate effect, I will look at patterns of population growth, housing size (square footage), and changes in heating fuel in addition to the heretofore studied differences between new and old residential buildings in temperature response.

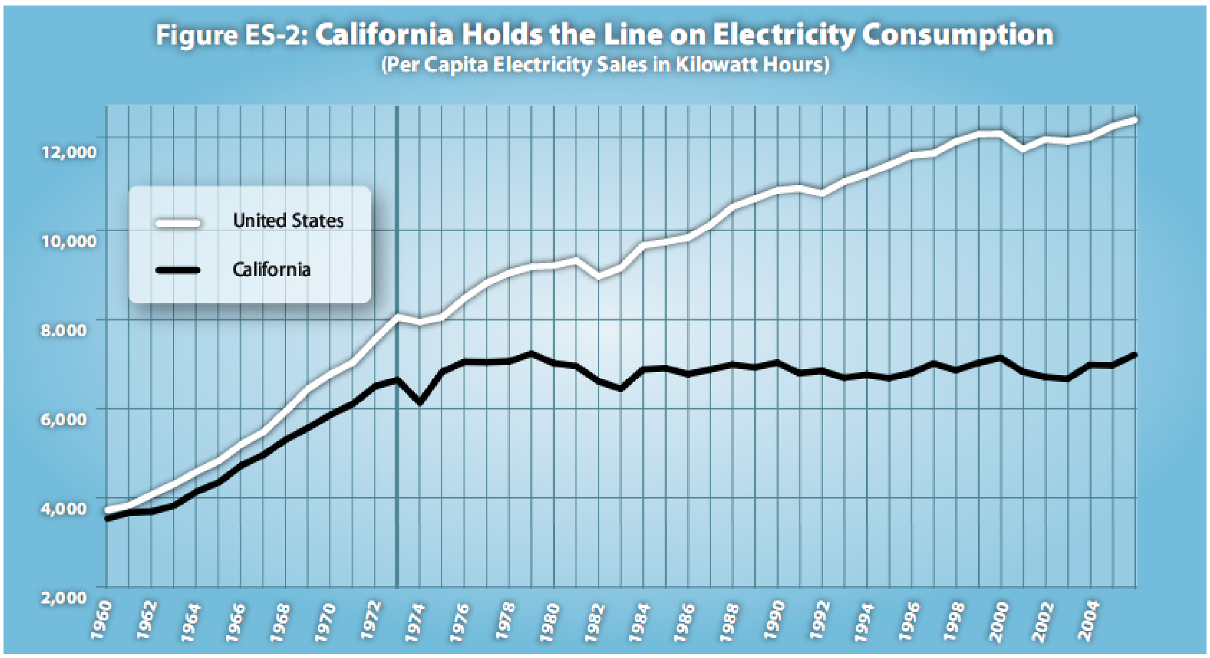
1.7 Conclusion

The contribution of this paper is to focus on the relationship between building vintage and temperature response in residential buildings in California. The main finding is that temperature response for buildings varies by vintage: new buildings (1970-2000) have a statistically significantly higher temperature response (*i.e.* use more electricity in response to higher temperature) than old buildings (pre-1970). This is robust to many specifications. The cumulative positive effect for temperature response in new buildings means that increased air conditioning ownership and other factors have outweighed other energy-saving impacts, such as building standards applied to new residential buildings.

This result has two main implications, one for electricity demand forecasting and one for climate change impacts. First, since new residential buildings have higher temperature response, this means that the average temperature response is expected to go up as new buildings are added. Peak electricity load will also increase, even with climate held constant. Second, if temperatures increase due to climate change, the new residential buildings will exacerbate the increase in peak load.

1.8 Figures and Tables

Figure 1.1: The "Rosenfeld" Curve. Per capita electricity sales for California and the United States, annually from 1960-2006. Source: California Energy Commission (2007).



Source: California Energy Commission.

Figure 1.2: Per capita *residential* electricity sales for eight western states, 1963-2004 . Source: Energy Information Administration (2007).

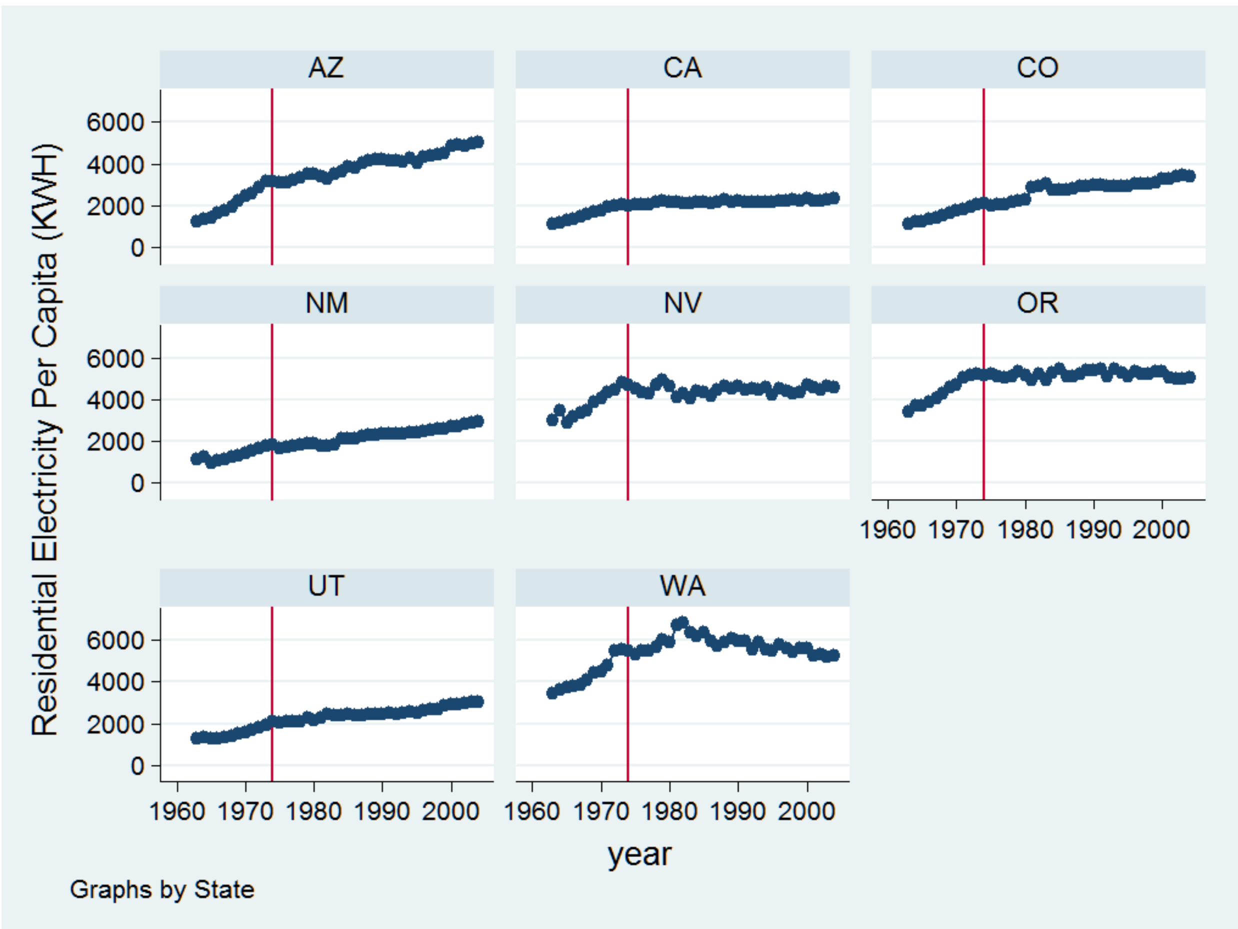


Figure 1.3: Variation in building vintage in Riverside County, California, USA. Shading represents proportion of buildings built since 1980. Darker means higher proportion of new buildings.

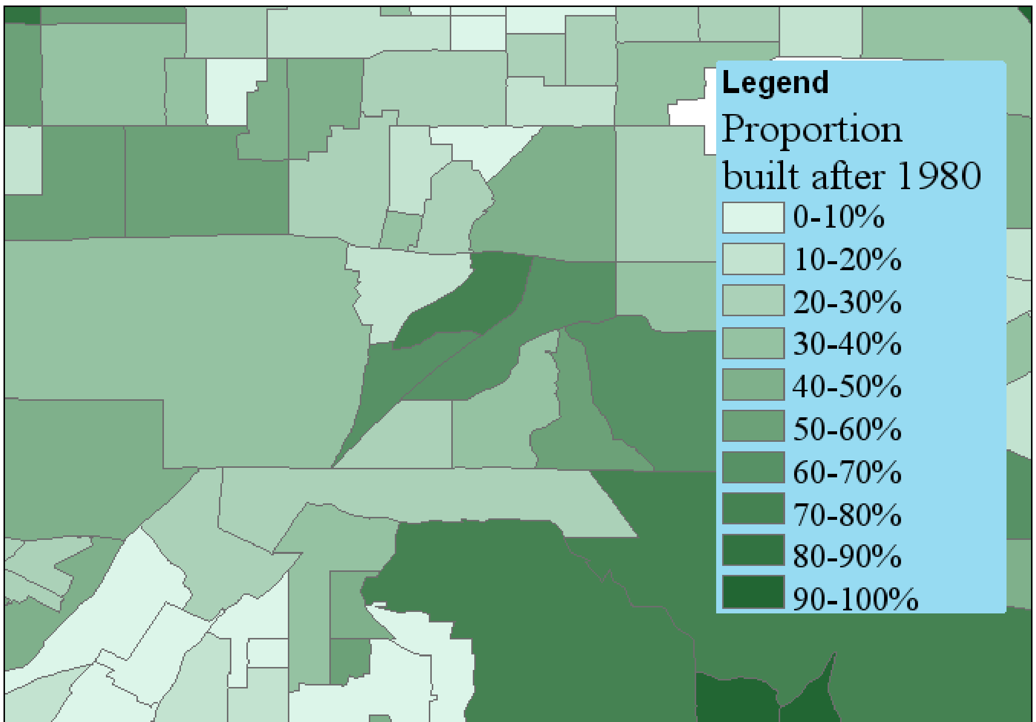


Table 1.1: Temperature bins.

Bin Number	Temperature Range
bin0	0-51.96°F
bin1	51.96-55.89°F
bin2	55.89-59.25°F
bin3	59.25-62.70°F
bin4	62.70-66.39°F
bin5	66.39-70.54°F
bin6	70.54-74.37°F
bin7	74.37-78.30°F
bin8	78.30-84.02°F
bin9	84.02-130°F

Figure 1.4: Electricity Use (KWH) vs time for one sample household. Source: Author's data.

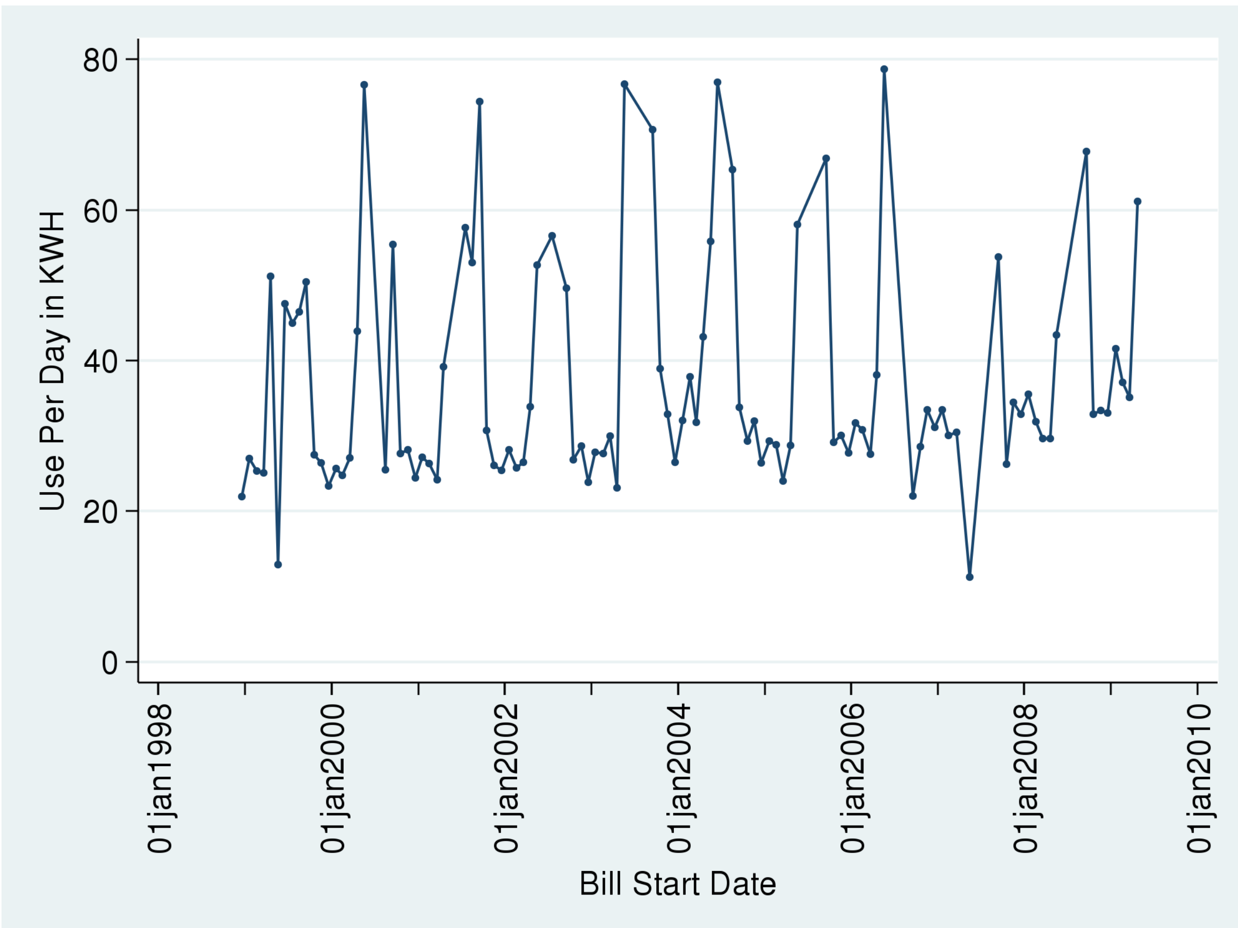
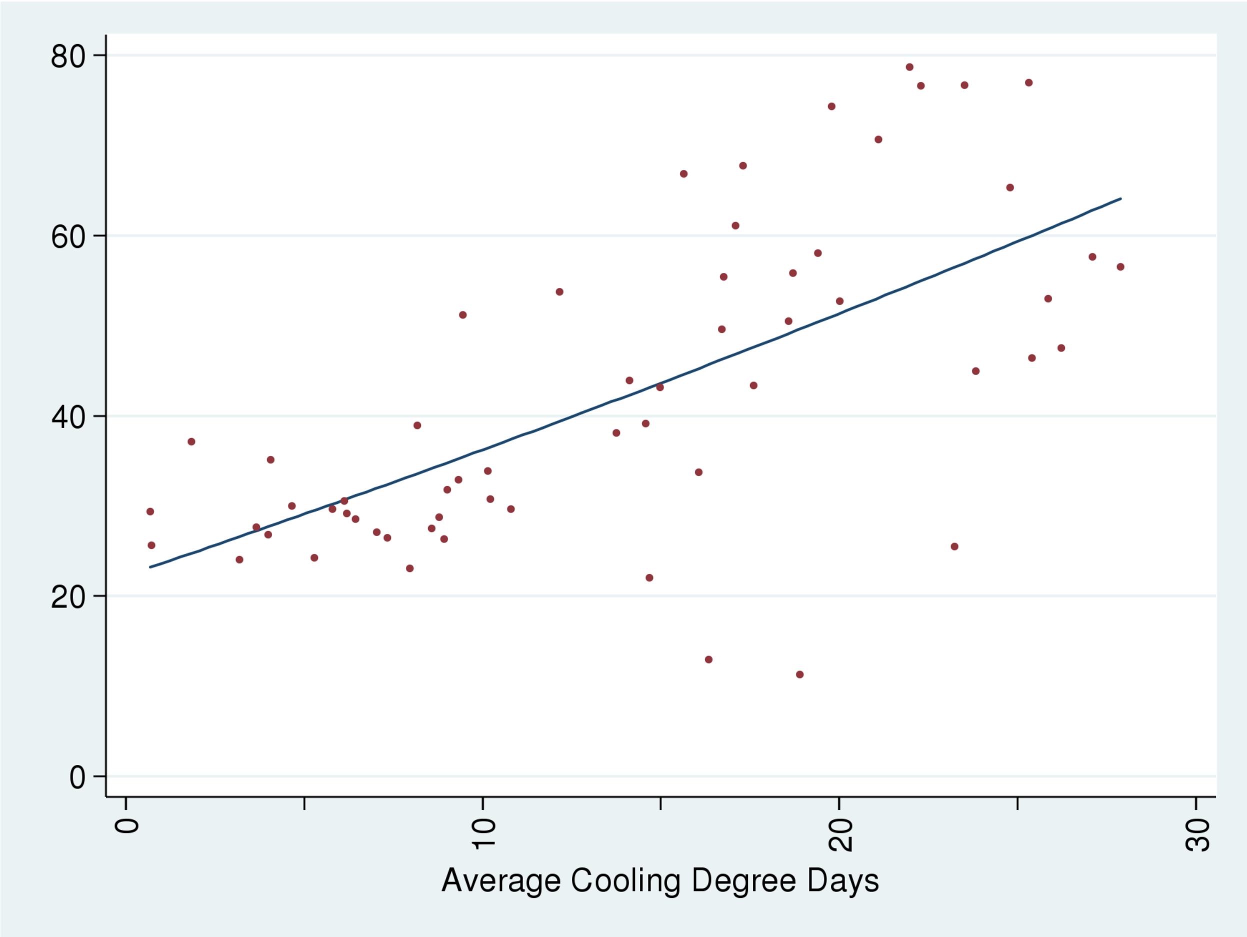


Figure 1.5: Electricity Use (KWH) vs cooling degree days for one sample household. Source: Author's data.



Note: Days with heating degree days were omitted.

Table 1.2: Summary Statistics.

Variable	Mean	Std. Dev.	Min	Max
BILLING DATA				
useperday (in kWh)	24.73	14.20	2.03	79.97
days	30.43	1.50	26	34
ASSESSOR'S DATA				
Building Age				
proportion built prior to 1970	0.15	0.36	0	1
proportion built in 1970s	0.16	0.36	0	1
proportion built in 1980s	0.54	0.50	0	1
proportion built in 1990s	0.15	0.36	0	1
Other Characteristics				
Square Feet	1750	480	360	7138
Has Central Air Conditioning?	0.88	0.28	0	1
for pre1970s	0.407			
for 1970s	0.847			
for 1980s	0.986			
for 1990s	0.995			
Observations (no subsampling)	5,106,398			

Table 1.3: Summary Statistics.

Variable	Mean	Std. Dev.	Min	Max
BILLING DATA				
useperday	21.61	14.56	2.03	79.97
days	30.43	1.52	26	34
CENSUS DATA				
Building Age				
proportion built prior to 1970	0.23	0.24	0	1
proportion built in 1970s	0.20	0.16	0	1
proportion built in 1980s	0.36	0.22	0	0.94
proportion built in 1990s	0.21	0.21	0	0.98
Type of Structure				
proportion SingleFamily	0.64	0.30	0	1
proportion MultiFamily	0.28	0.28	0	1.00
proportion MotorOther	0.073	0.15	0	0.84
Other Characteristics				
Average Bedrooms	2.57	0.62	0.89	4.36
Average Rooms	5.23	1.01	2.31	8.00
Average Income	$48,200	$16,600	$13,000	$108,900
Observations (1-in-5 subsample)	5303019			

Table 1.4: Estimation results, temperature response with CDD and HDD parameterization, assessor's data.
Dependent variable is $ln(KWH\_perday)$

VARIABLES	(A1)	(A2)	(A3)	(A4)
CDD	0.0553*** [0.000330]	0.0460*** [0.000930]	0.0355*** [0.00149]	0.0476*** [0.00321]
HDD	0.0274*** [0.000268]	0.0293*** [0.000905]	0.0231*** [0.00132]	0.0326*** [0.00286]
CDD^2	-0.000283*** [1.61e-05]	-0.000185*** [4.44e-05]	0.000234*** [8.27e-05]	-0.000270* [0.000164]
HDD^2	-0.000590*** [1.87e-05]	-0.000743*** [6.64e-05]	-0.000260*** [8.93e-05]	-0.000802*** [0.000191]
Built in 1990×CDD		0.0149*** [0.00121]	0.0108*** [0.00160]	0.00871*** [0.00300]
Built in 1980×CDD		0.0117*** [0.00103]	0.00482*** [0.00141]	0.00827*** [0.00246]
Built in 1970×CDD		0.00393* [0.00201]	0.000171 [0.00210]	0.000292 [0.00371]
Built in 1990×HDD		-0.00049 [0.00109]	-0.00306** [0.00139]	-0.00364 [0.00248]
Built in 1980×HDD		-0.00195** [0.000976]	-0.00502*** [0.00125]	-0.00359* [0.00208]
Built in 1970×HDD		-0.00800*** [0.00161]	-0.00980*** [0.00169]	-0.00984*** [0.00295]
Built in 1990× CDD^2		-0.000242*** [5.61e-05]	-0.000151** [6.81e-05]	-3.00E-05 [0.000116]
Built in 1980× CDD^2		-0.000130*** [5.00e-05]	7.86E-05 [5.98e-05]	-2.47E-05 [9.66e-05]
Built in 1970× CDD^2		7.56E-05 [0.000114]	0.000147 [0.000113]	0.000384** [0.000195]
Built in 1990× HDD^2		0.000128* [7.71e-05]	0.000114 [9.32e-05]	0.000201 [0.000154]
Built in 1980× HDD^2		0.000156** [7.02e-05]	0.000263*** [8.61e-05]	0.000261* [0.000133]
Built in 1970× HDD^2		0.000674*** [0.000106]	0.000737*** [0.000109]	0.000598*** [0.000186]

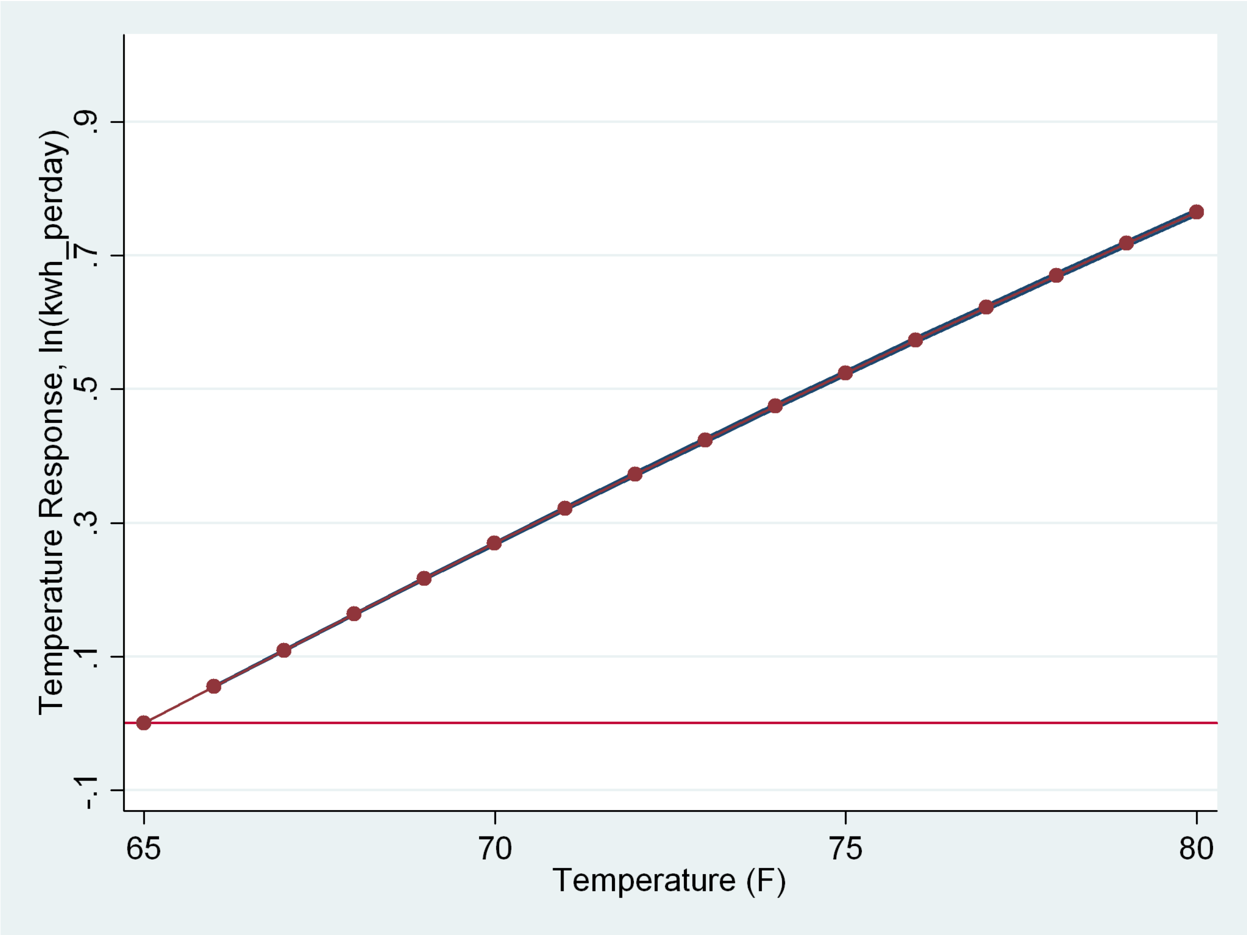
regression results continued

continuation of regression results

VARIABLES	(A1)	(A2)	(A3)	(A4)
Central Air Conditioning×CDD			0.0160*** [0.00186]	0.00803** [0.00318]
Central Air Conditioning×HDD			0.00822*** [0.00164]	0.000798 [0.00275]
Central Air Conditioning× CDD^2			-0.000554*** [8.98e-05]	-0.000287* [0.000151]
Central Air Conditioning× HDD^2			-0.000534*** [0.000110]	-0.00014 [0.000181]
Square Feet×CDD			-0.00516*** [0.000482]	0.00621** [0.00295]
Square Feet×HDD			-0.00239*** [0.000360]	0.00540** [0.00257]
Square Feet× CDD^2			0.000241*** [2.66e-05]	-0.000235* [0.000134]
Square Feet× HDD^2			0.000243*** [2.16e-05]	-4.91E-05 [0.000164]
Constant	2.704*** [0.00122]	2.705*** [0.00138]	2.710*** [0.00151]	2.597*** [0.00244]
Observations	5,625,517	5,625,517	5,625,517	1,652,525
R-squared	0.363	0.366	0.367	0.414
Number of households	118,252	118,252	118,252	37,984

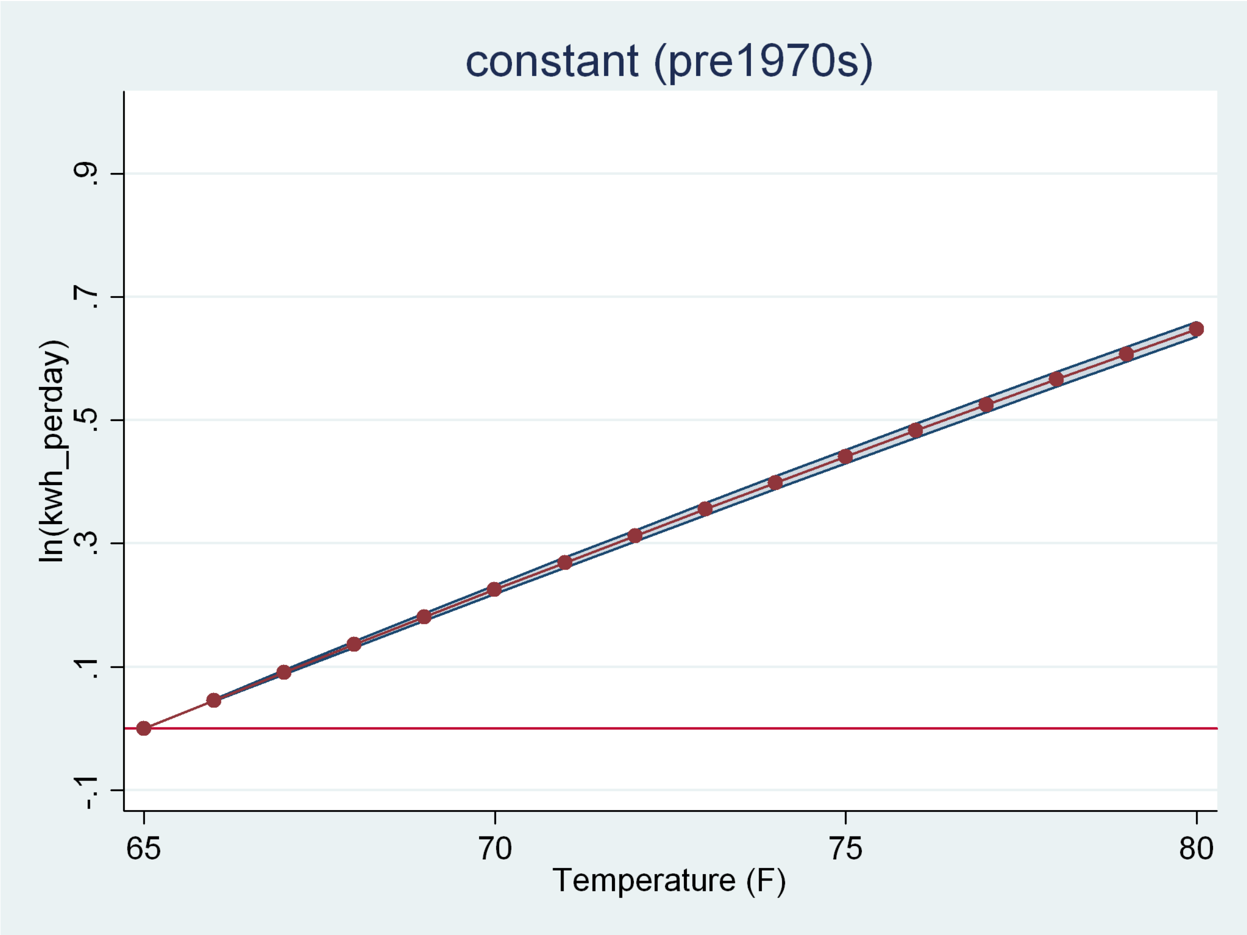
Includes household-level fixed effects. *,**,*** represent 10%, 5%, and 1% statistical significance, respectively. Robust standard errors clustered at the Zip9-level. † notes that the square feet variable has been demeaned (1750 square feet) and rescaled by the population standard deviation (480 square feet).

Figure 1.6: Estimation results, temperature response with CDD and HDD parameterization, assessor's data, average across all vintages
Dependent variable is $ln(KWH\_perday)$



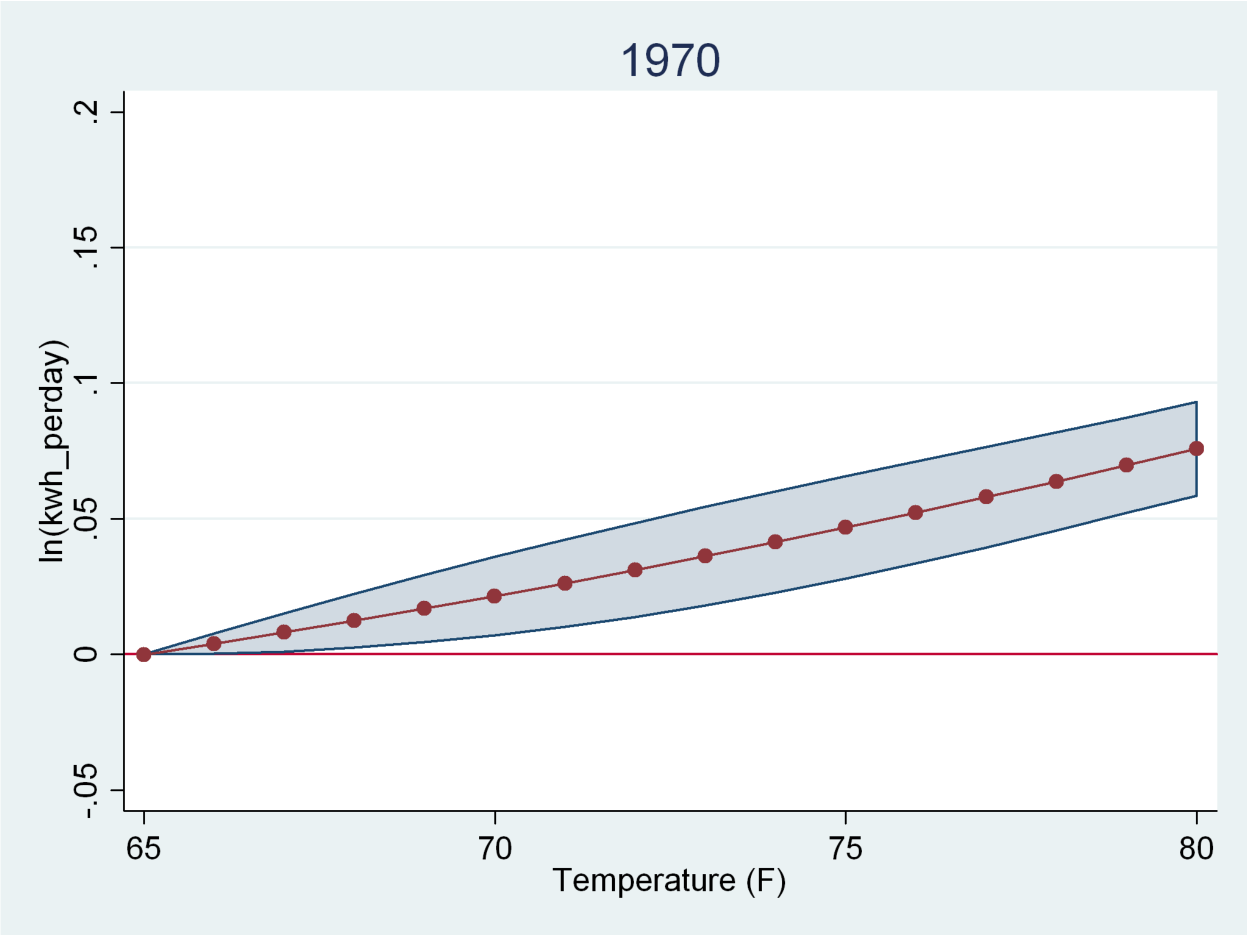
The range represents the 95% confidence interval.

Figure 1.7: Estimation results, temperature response with CDD and HDD parameterization, assessor's data, by vintage. pre1970s reference curve.
Dependent variable is $ln(KWH\_perday)$



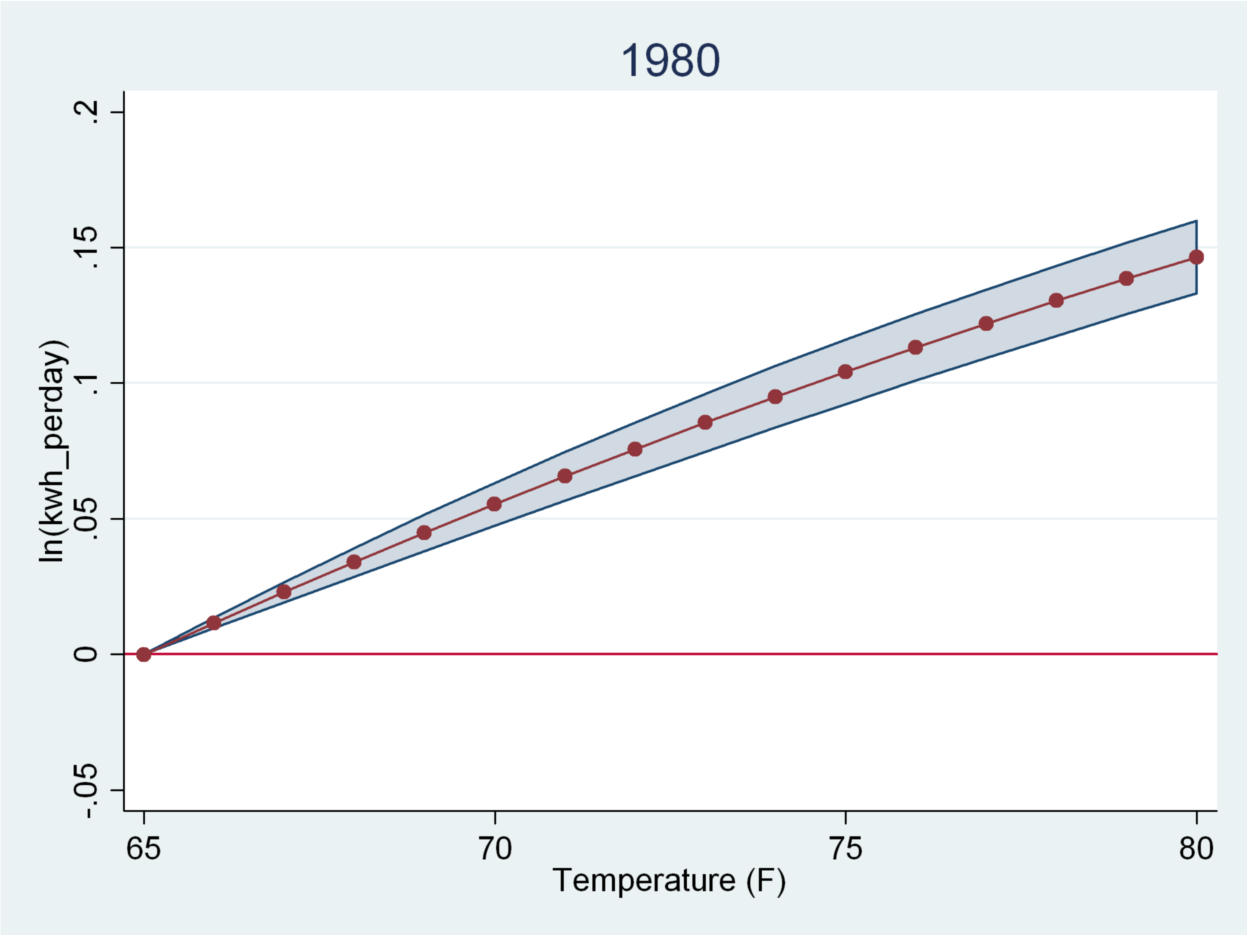
The range represents the 95% confidence interval with robust standard errors.

Figure 1.8: Estimation results, temperature response with CDD and HDD parameterization, assessor's data, by vintage. 1970s relative to pre1970s curve.
Dependent variable is $ln(KWH\_perday)$



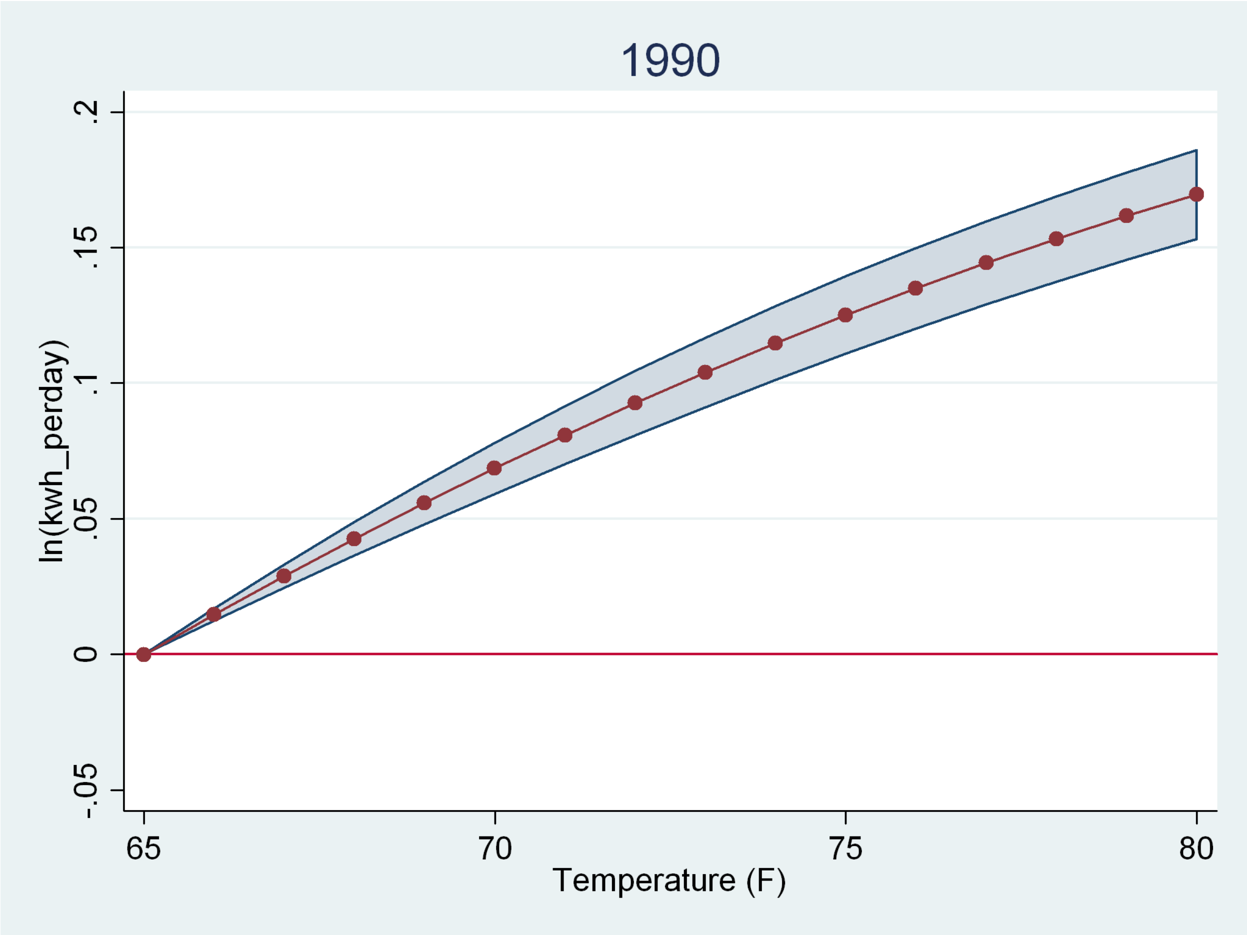
The range represents the 95% confidence interval with robust standard errors.

Figure 1.9: Estimation results, temperature response with CDD and HDD parameterization, assessor's data, by vintage. 1980s relative to pre1970s curve.
Dependent variable is $ln(KWH\_perday)$



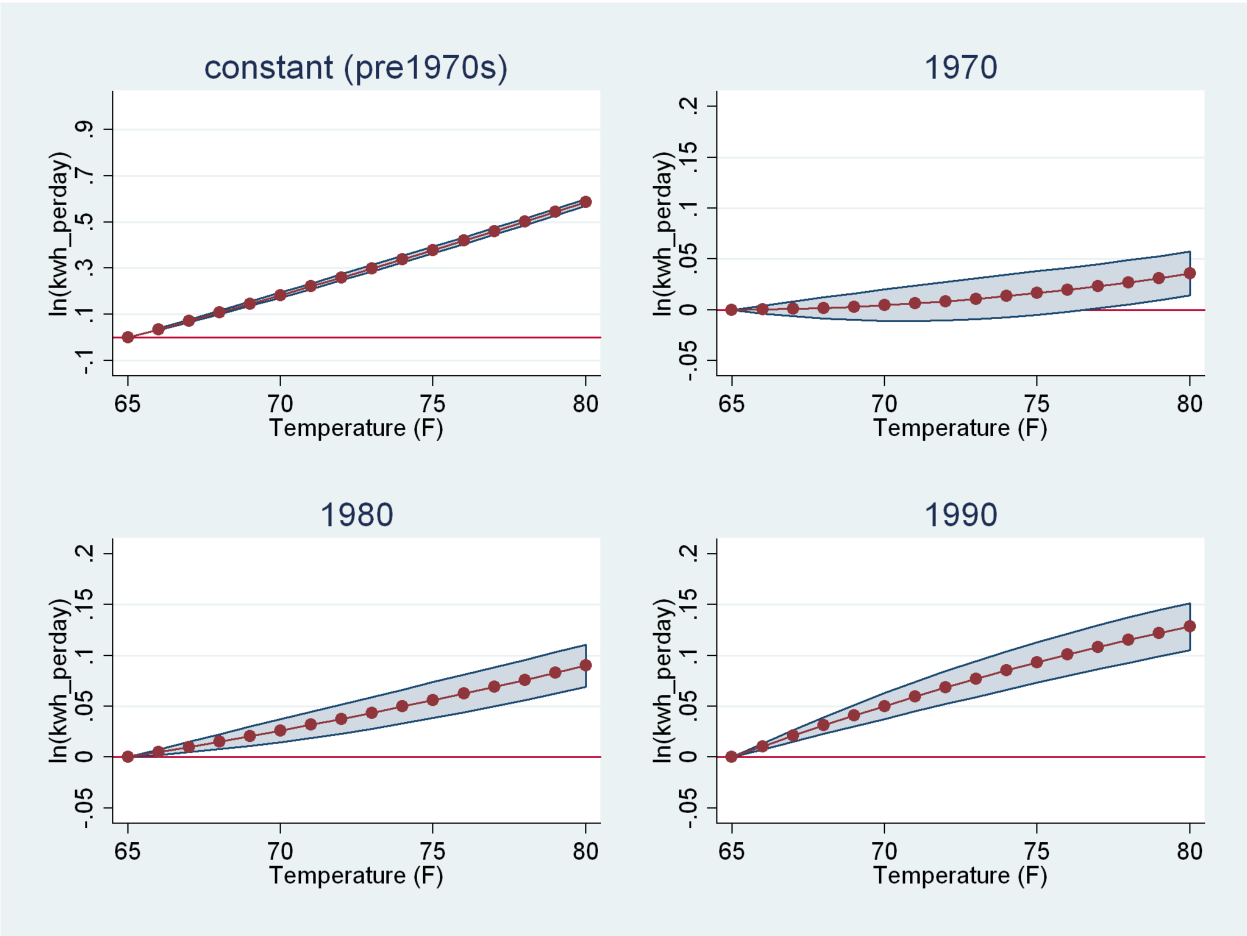
The range represents the 95% confidence interval with robust standard errors.

Figure 1.10: Estimation results, temperature response with CDD and HDD parameterization, assessor's data, by vintage. 1990s relative to pre1970s curve.
Dependent variable is $ln(KWH\_perday)$



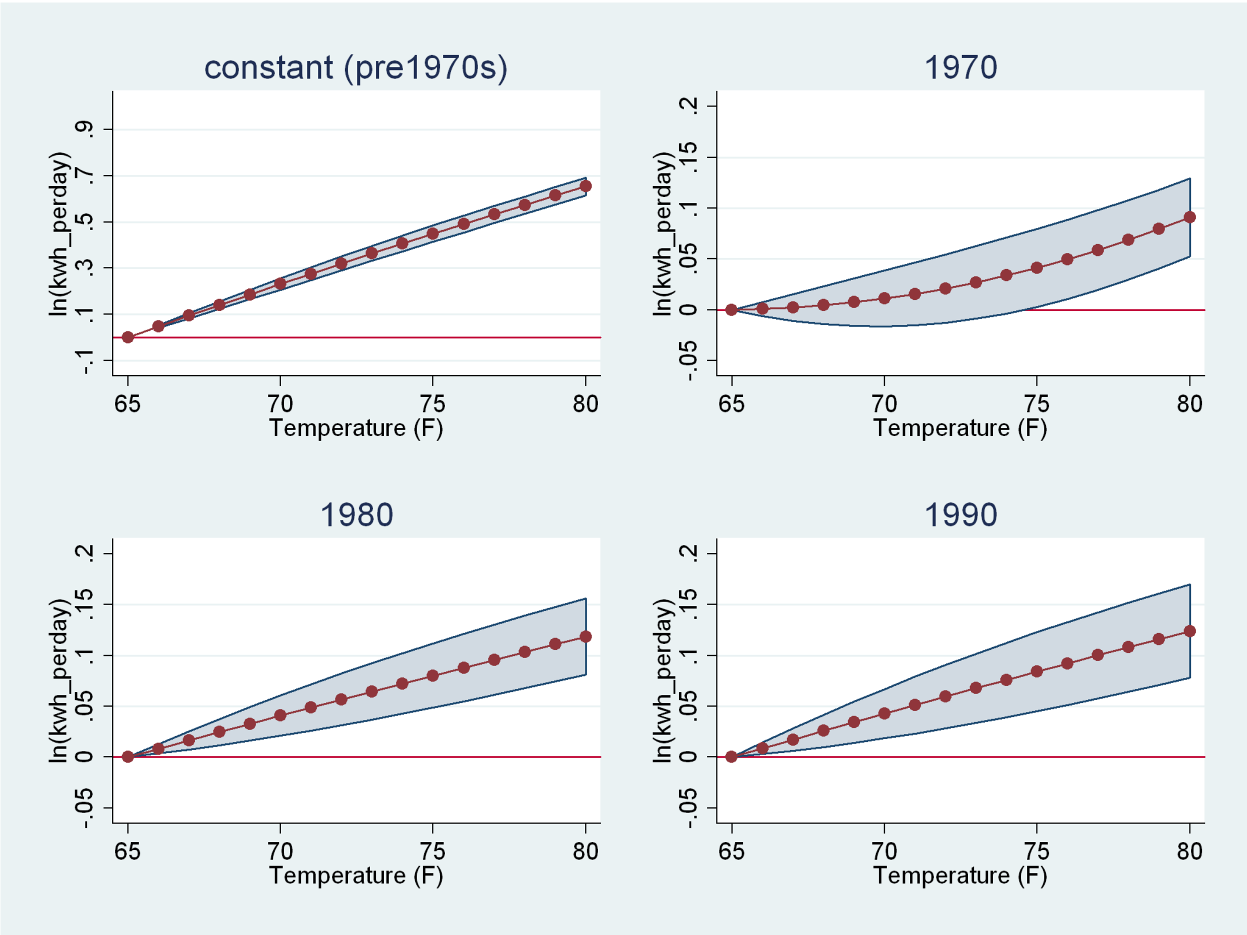
The range represents the 95% confidence interval with robust standard errors.

Figure 1.11: Estimation results, temperature response with CDD and HDD parameterization, assessor's data, by vintage, with controls.
Dependent variable is $ln(KWH\_perday)$



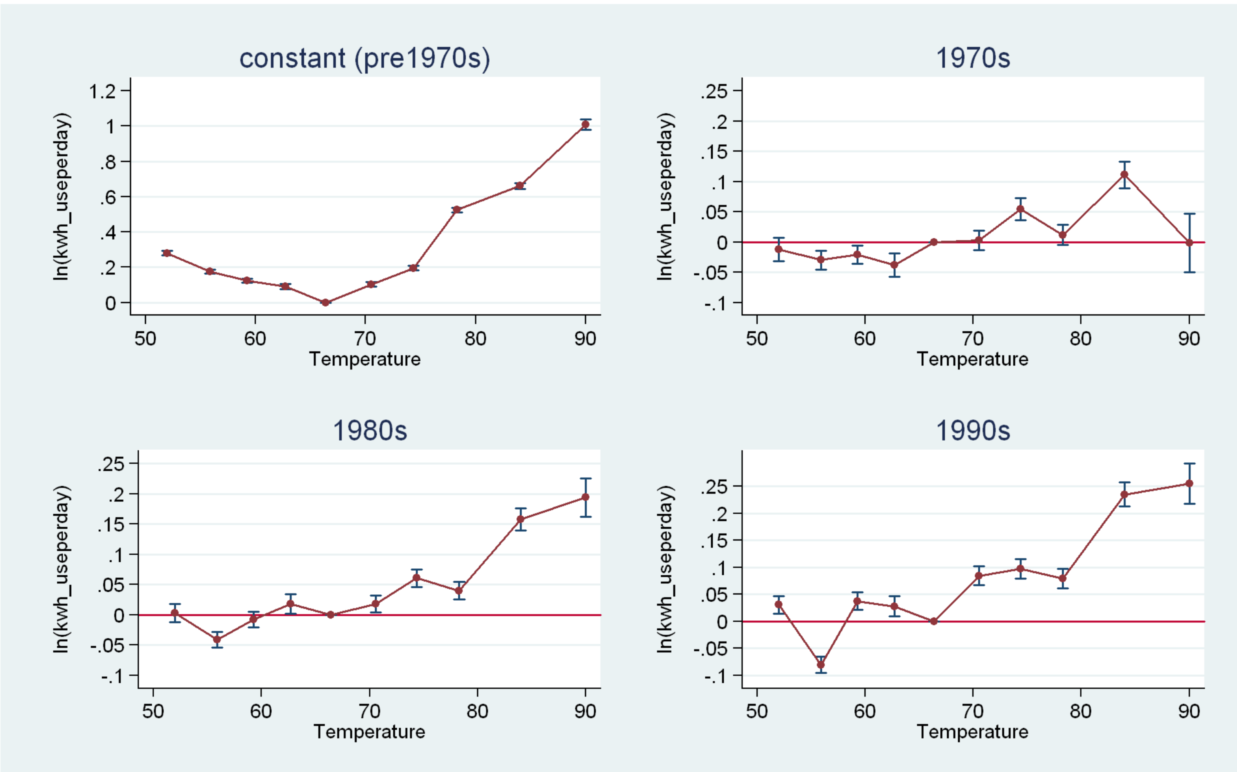
The range represents the 95% confidence interval with robust standard errors. Top left graph is the reference curve for pre1970s buildings. The remaining curves are the relative temperature responses of the other vintages.

Figure 1.12: Estimation results, temperature response with CDD and HDD parameterization, assessor's data, by vintage, with controls. Home size restricted to 1300-1600sqft. Dependent variable is $ln(KWH\_perday)$



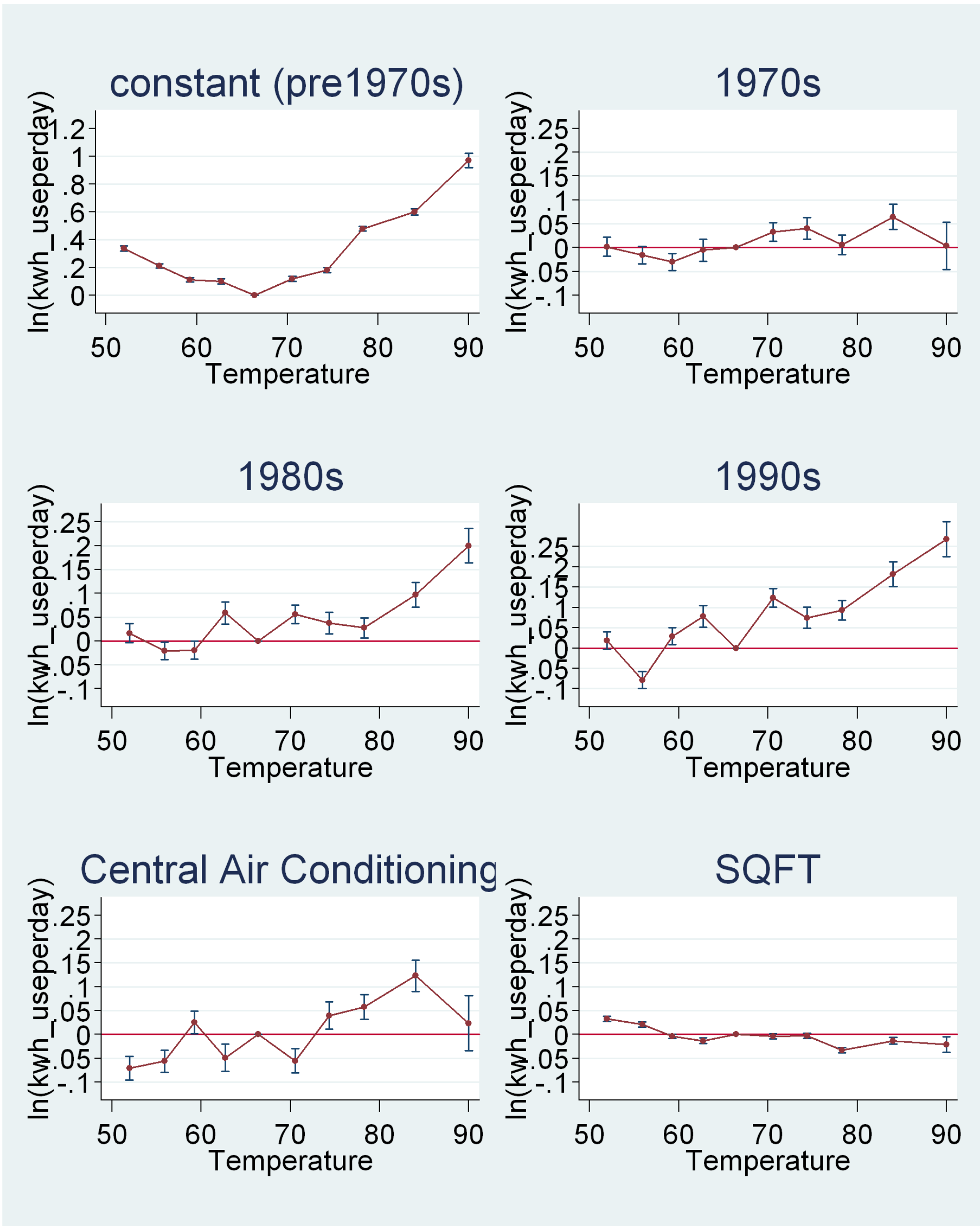
The range represents the 95% confidence interval with robust standard errors. Top left graph is the reference curve for pre1970s buildings. The remaining curves are the relative temperature responses of the other vintages.

Figure 1.13: Estimation results, temperature response with binning, assessor's data, by vintage, no controls.
Dependent variable is $ln(KWH\_perday)$



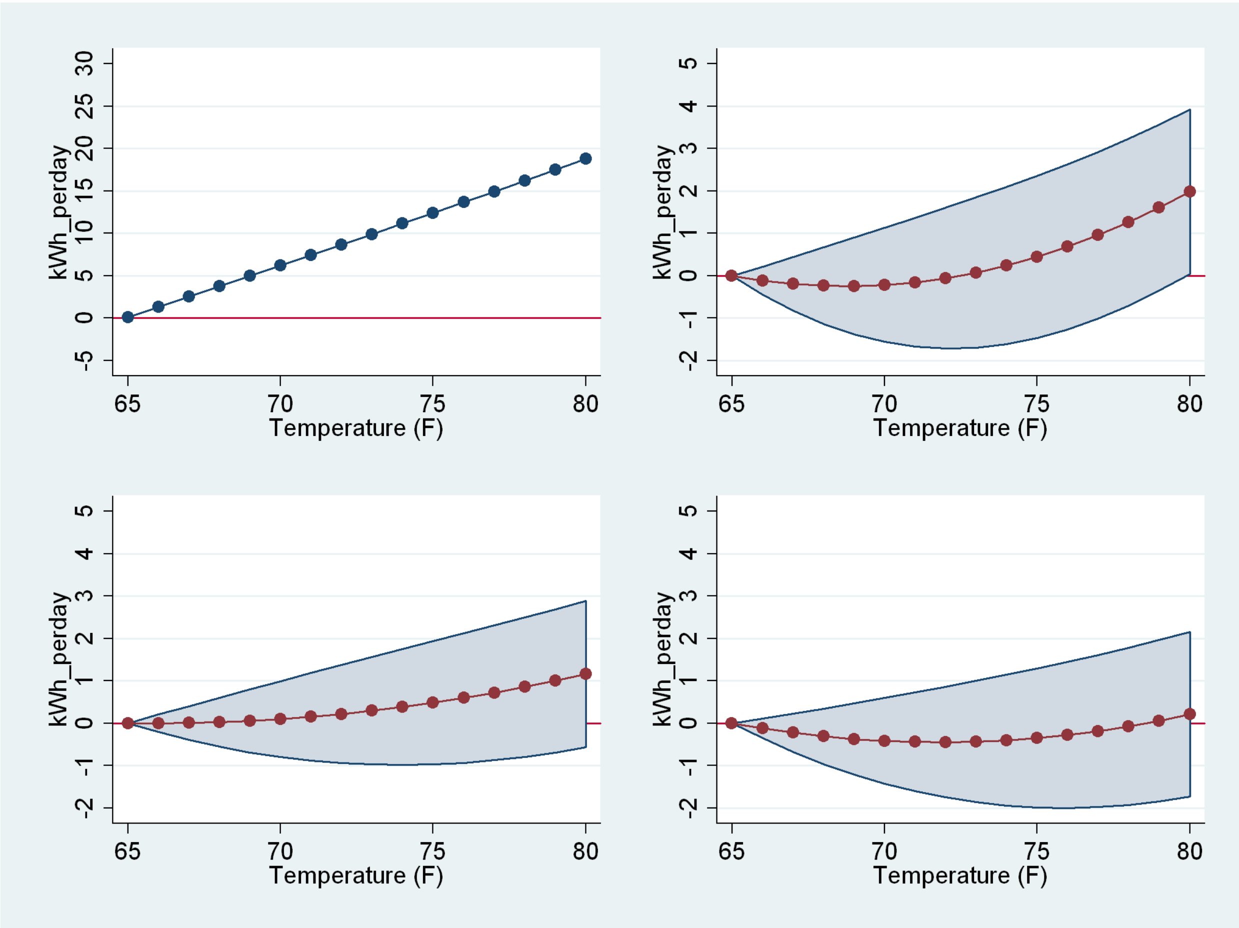
The range represents the 95% confidence interval with robust standard errors. Top left graph is the reference curve for pre1970s buildings. The remaining curves are the relative temperature responses of the other vintages.

Figure 1.14: Estimation results, temperature response with binning, assessor's data, by vintage, with controls.
Dependent variable is $ln(KWH\_perday)$



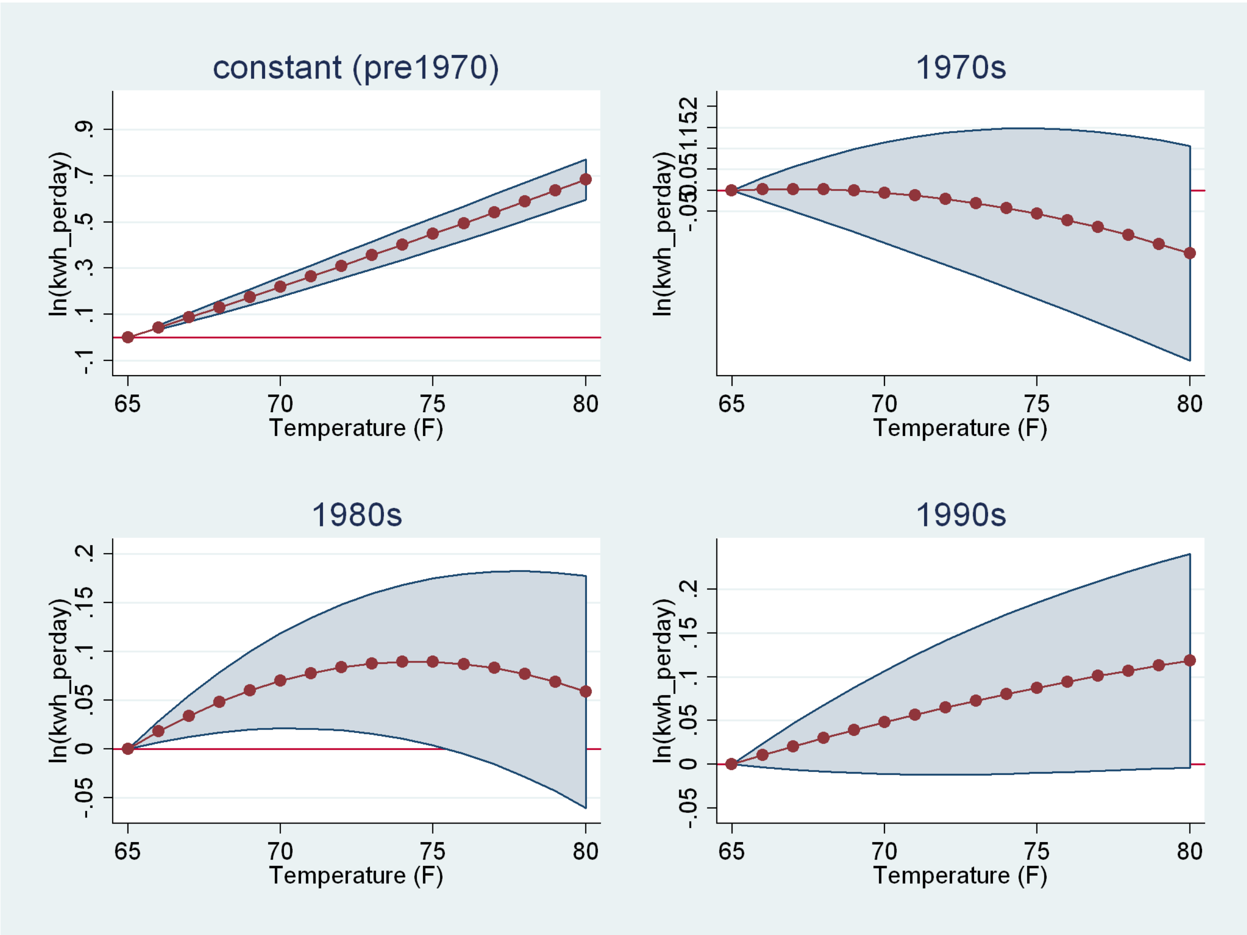
The range represents the 95% confidence interval with robust standard errors. Top left graph is the reference curve for pre1970s buildings. The next three curves are the relative temperature responses of the other vintages. The bottom curves plot the impact of central air conditioning and square footage.

Figure 1.15: Estimation results, temperature response with CDD and HDD parameterization, assessor's data, by vintage, with controls. Home size restricted to 1300-1600sqft. Dependent variable is KWH_perday



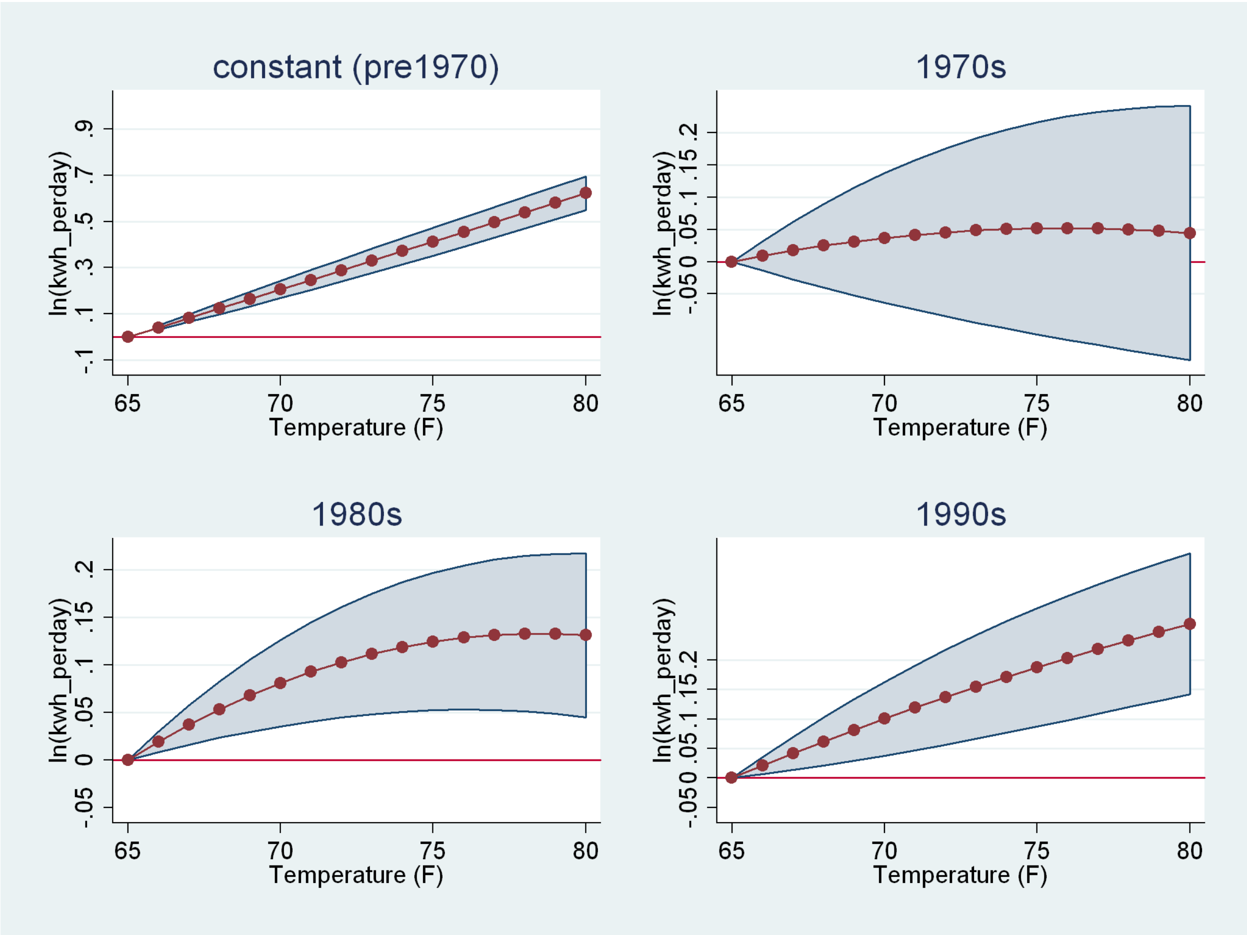
The range represents the 95% confidence interval with robust standard errors. Top left graph is the reference curve for pre1970s buildings. The remaining curves are the relative temperature responses of the other vintages.

Figure 1.16: Estimation results, temperature response with CDD and HDD parameterization, census data, by vintage, no controls.
Dependent variable is $ln(KWH\_perday)$



The range represents the 95% confidence interval with robust standard errors. Top left graph is the reference curve for pre1970s buildings. The next three curves are the relative temperature responses of the other vintages.

Figure 1.17: Estimation results, temperature response with CDD and HDD parameterization, census data, by vintage, with controls for type of structure, bedrooms, and income. Dependent variable is $ln(KWH\_perday)$



The range represents the 95% confidence interval with robust standard errors. Top left graph is the reference curve for pre1970s buildings. The next three curves are the relative temperature responses of the other vintages. Variation in temperature response by the three controls (structure, bedrooms, and income) are omitted.

Table 1.5: Estimation results, Differences across vintage for total usage, assessor's data
Dependent variable is $ln(KWH\_perday)$

VARIABLES	(T1)	(T2)	(T3)
Built in1990s	0.145***	-0.156***	-0.196***
	[0.0124]	[0.0137]	[0.0149]
Built in1980s	0.0459***	-0.126***	-0.133***
	[0.00959]	[0.0120]	[0.0130]
Built in1970s	0.177***	0.0118	0.0445***
	[0.0139]	[0.0133]	[0.0146]
Square Feet†		0.158***	0.189***
		[0.00341]	[0.00381]
Central Air Conditioning		0.0980***	-0.00343
		[0.0147]	[0.0157]
Constant	2.833***	2.928***	2.681***
	[0.00858]	[0.00963]	[0.0108]
random effects	yes	yes	yes
controls for Temperature Response	no	no	yes
Observations	5,625,517	5,625,517	5,625,517
Number of aididlong	118,252	118,252	118,252
Robust standard errors in brackets			
*** p¡0.01, ** p¡0.05, * p¡0.1			

*,**,*** represent 10%, 5%, and 1% statistical significance, respectively. Robust standard errors clustered at the Zip9-level. Temperature response controls include square feet, central air conditioning, and vintage dummies interacted with the quadratic degree day parameterization. †The square feet variable has been demeaned and rescaled by the population standard deviation.

Figure 1.18: Simulation of Riverside average temperature response in 2020, with and without new building stock. Source: Author's calculations.

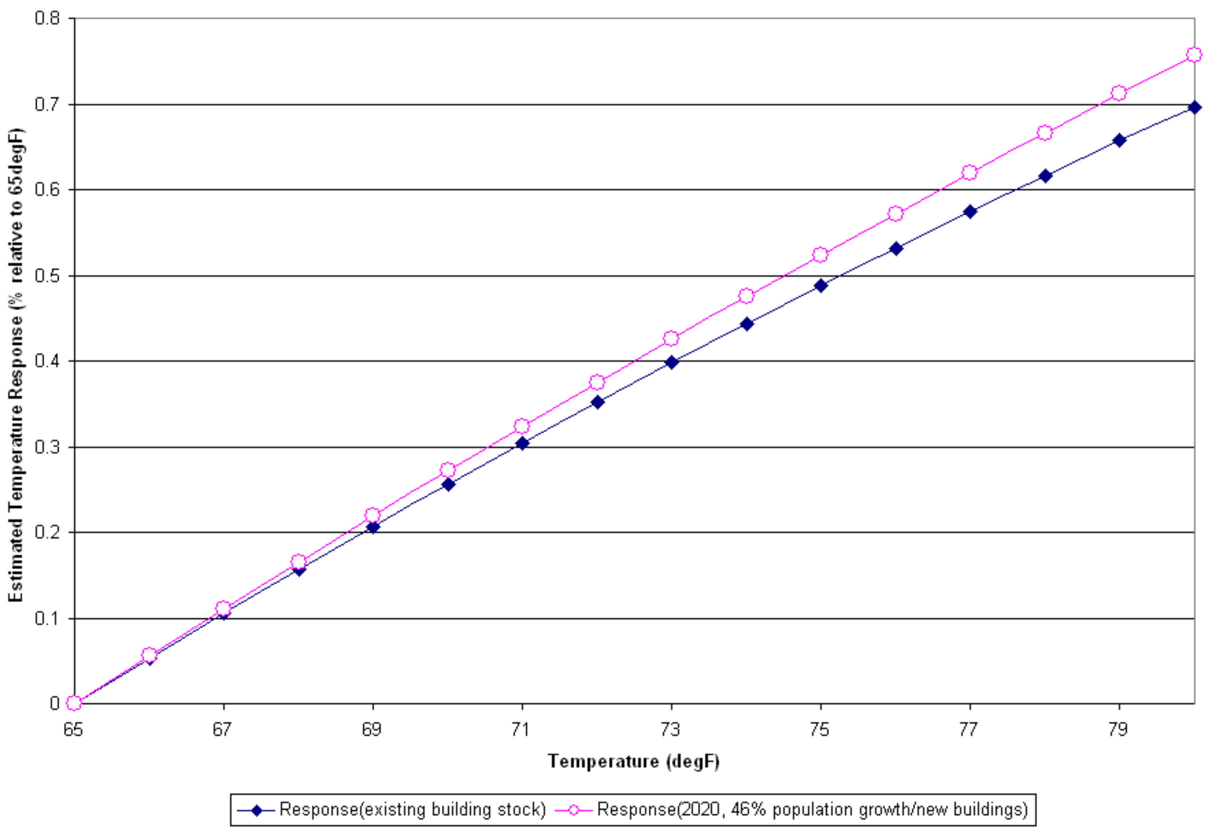


Table 1.6: Comparison of Air Conditioning Saturation by Climate Zone for Old and New Homes. SOURCE: RASS 2004

		Central Air		Central or Room Air	
Zone	Geography	1990s	pre1970s	1990s	pre1970s
1	Inland	56%	23%	63%	37%
2	Inland	96%	55%	97%	78%
3	Inland	93%	61%	95%	79%
4	Coastal	69%	30%	72%	41%
5	Coastal	27%	4%	29%	8%
7	Inland	93%	59%	93%	73%
8	Coastal	77%	21%	80%	32%
9	Inland	84%	39%	85%	59%
10*	Inland	94%	53%	96%	76%
11	Coastal	60%	12%	68%	25%
12	Coastal	75%	51%	82%	81%
13	Coastal	68%	22%	69%	32%

Note: Zone refers to Forecast Climate Zones as determined by the California Energy Commission. Zone 10 includes Riverside County. A map of the zones is given below in Figure 1.19.

Figure 1.19: California Energy Commission Forecast Climate Zones. Source: California Energy Commission (2007), page 24.

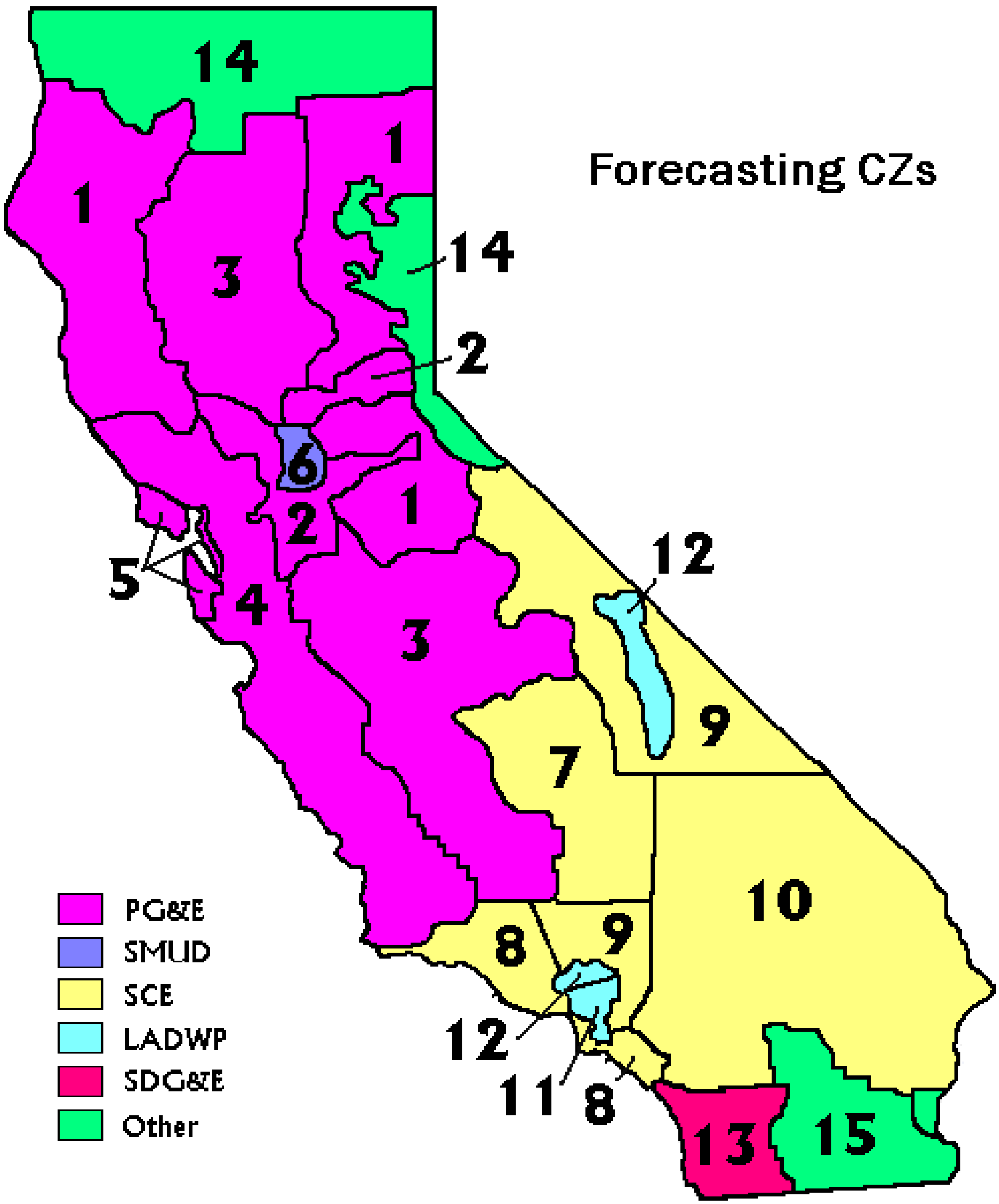


Table 1.7: Comparison of Air Conditioning Saturation by Vintage in Forecast Climate Zone 10. SOURCE: RASS 2004

Vintage	Central Air	Room Air	Central or Room
pre1970s	53%	22%	76%
1970	80%	9%	88%
1980	89%	2%	91%
1990	94%	1%	96%

Note: Forecast Climate Zone 10 includes Riverside County.

Table 1.8: Savings from Building Standards in 2005 and the Implied Reduction In Temperature Response. SOURCE: California Energy Commission reports, author's calculations

Standard	Estimated Savings (GWH)	Percent of Total Load	Population Increase Since Standard	Implied Impact on Temperature Response (Cumulative Standards)
Building Standard1992	310.7	0.4%	15%	-34 to -56%
Building Standard1984	1074.8	1.3%	29%	-31 to -53%
Building Standard1979	878.7	1.1%	36%	-25 to -45%
Building Standard1975	3166.9	3.8%	41%	-20 to -38%

Notes: These values include only the top 5 utilities: PG&E, SDG&E, SCE, LADWP, and SMUD. These utilities supply electricity to a wide majority of the state's population. Total residential load is 83600 GWH for these utilities. To interpret columns 4 and 5 in the second to last row, 36% = (change in population from 1979 to 2005)/(population in 2005) and -25 to -45% is the implied percent reduction from all standards prior to and including the 1979 standards. The calculation range is given by a low and high assumption, 0.1 and 0.25, of the proportion of load that is temperature response. An adjustment factor of 50% was also used to crudely account for the fact that growth has been faster in hotter inland areas and that new homes are larger. These numbers should be treated as speculative because the details of how the estimated savings were calculated are not fully known beyond that which is described in the two CEC reports referenced in the text of my paper.

Table 1.9: Proportion of homes with central air conditioning on by vintage and time of day.

	Morning	Day	Evening	Night
pre1970s	0.49	0.71	0.73	0.46
1970	0.54	0.69	0.76	0.5
1980	0.51	0.66	0.76	0.47
1990	0.52	0.72	0.82	0.52

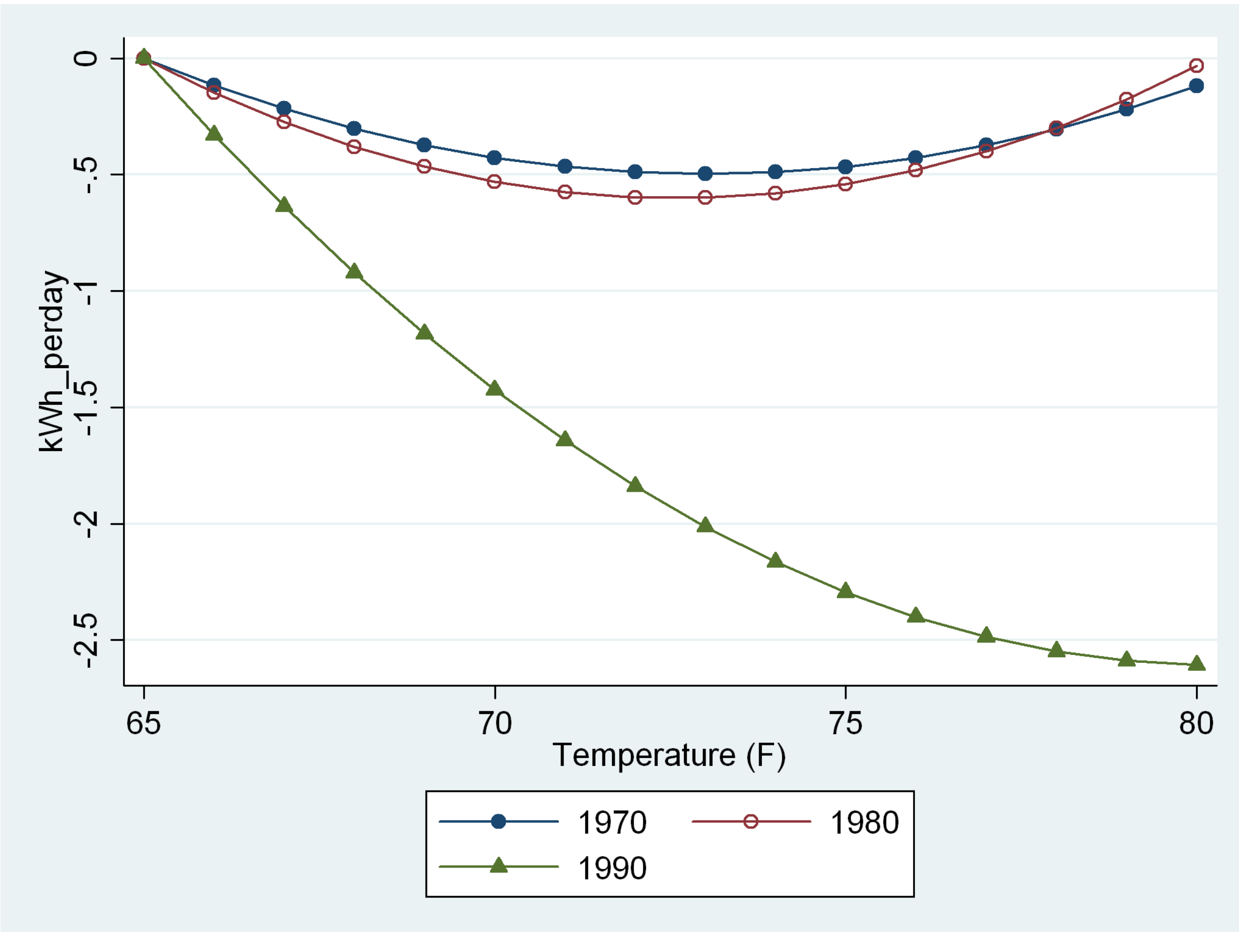
N is about 300. Sampling error is about 0.03 for each cell

Table 1.10: Average thermostat set points (in °F) conditional on central air conditioning on by vintage and time of day.

	Morning	Day	Evening	Night
pre1970s	0.49	0.71	0.73	0.46
1970	0.54	0.69	0.76	0.5
1980	0.51	0.66	0.76	0.47
1990	0.52	0.72	0.82	0.52

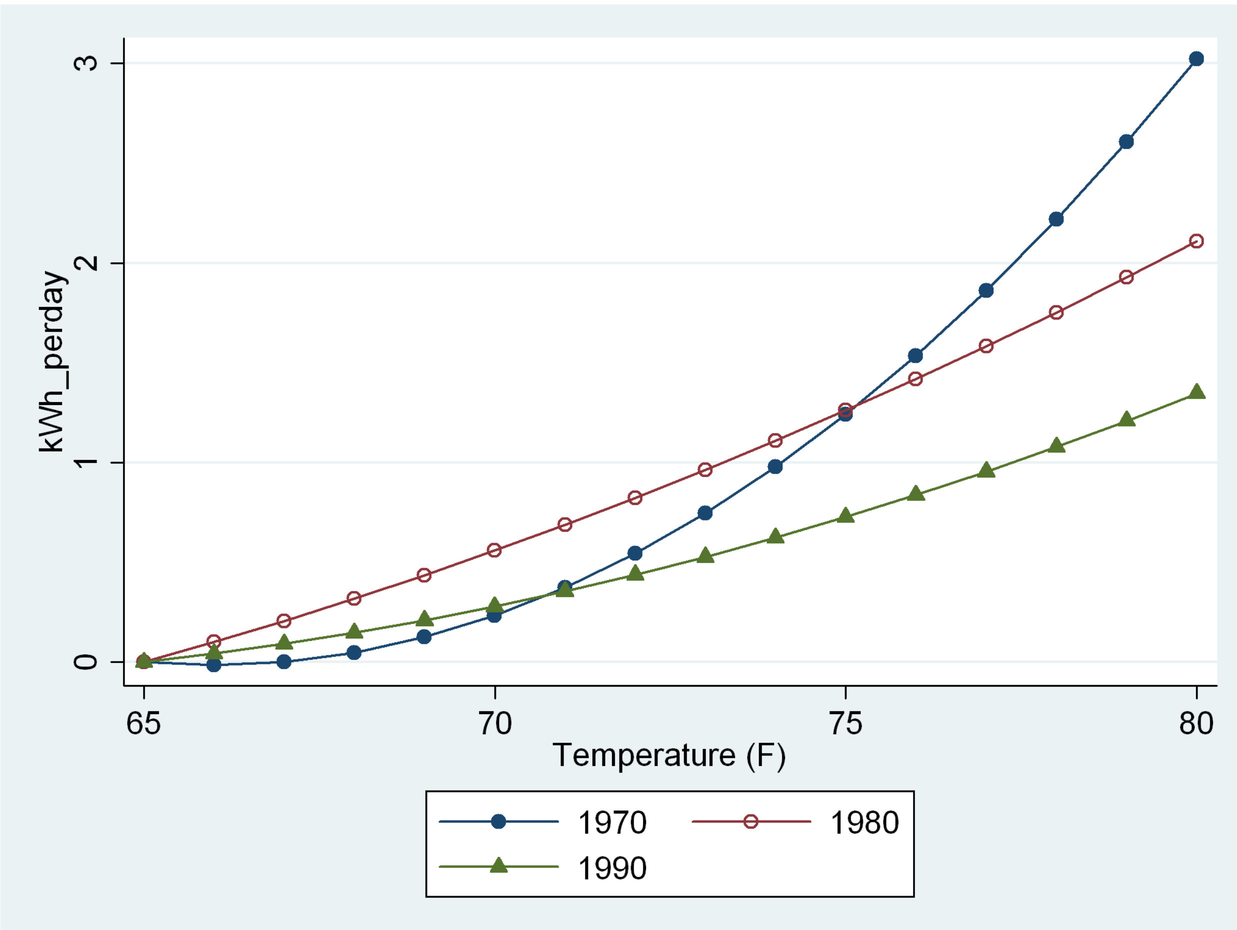
Standard deviation is about 3degF for each cell

Figure 1.20: Estimation results, difference in temperature response with CDD and HDD parameterization, assessor's data, by vintage.
Dependent variable is KWH_perday



This uses the assumption that $f(size) = sqft$.

Figure 1.21: Estimation results, difference in temperature response with CDD and HDD parameterization, assessor's data, by vintage, restricted to sqft in [1300,1600].
Dependent variable is KWH_perday



This uses the assumption that $f(size) = sqft$, but for a narrow range of sqft.

Chapter 2

Profiting from Regulation: An Event Study of the European Carbon Market

Co-authored with Erin Mansur and James Bushnell.

2.1 Introduction

There is a long-standing perception of a fundamental conflict between the interests of business and environmental regulators. In many cases regulators apply policies that increase production costs, restrict production, or otherwise constrain the actions of firms. There is a rich literature chronicling the impacts that regulations such as the Clean Air Act have had on industrial activity.[1] With greenhouse gas regulation on the horizon in the US and already under way in the European Union, the question of the impacts of these regulations on industry has taken center stage. As countries and regions around the world develop policies for limiting greenhouse gas (GHG) emissions, there is an understandably great interest in how these policies will impact the competitiveness, productivity, and profitability of the industries to which they are applied.

Measuring the economic impacts of GHG regulations obviously has direct relevance to setting the levels and timings of the regulations. Even setting aside the specific goals for GHG reductions, information about the overall magnitude and distribution of economic impacts has importance for the policy-making process. This is most starkly true in the case of cap-and-trade mechanisms, which create valuable new property rights in the form of emissions allowances or permits. These permits constitute the "currency" of cap and trade markets. They also provide an important tool to policy makers for distributing the revenues collected

[1] Gray (1987), Becker and Henderson (2000), Gray and Shadbegian (1998), List, Millimet and McHone (2004)

by the carbon regulation. The process of allocating emissions allowaances, while inevitably containing a strong element of political maneuvering, is usually grounded in a desire to offset some of the cost impacts of the introduction of carbon regulation. Industries that claim to bear the brunt of the abatement costs usually stake the largest claim to allocations of allowances.

However, for most industrial enterprises, changes in direct abatement costs are only one piece of a complicated profitability puzzle. The introduction of a price of CO_2 into an economy can have indirect impacts on firms that are not large CO_2 emitters. In most industries, increases in CO_2 costs will be reflected in output prices, and therefore revenues, as well as in costs. A more complete picture of these net impacts is necessary in any attempt to align allocations to the true economic impacts of CO_2 regulation on firms.

Indeed, the impact of regulations on profitability is ambiguous, even when those regulations have a substantial impact of costs. There are several mechanisms, ranging from restricting entry (e.g. Ryan (2005)) to raising rivals' costs (e.g. Puller (2006)) through which revenue increases can outstrip cost increases, enhancing profitability.[2] With cap-and-trade regulations, the free allocation of emissions allowances adds an additional source of revenue. In the case of GHG markets, these assets can total hundreds of billions of dollars.

Despite the politically motivated tendency to award emissions allowances proportionally to emissions, several papers have concluded that this likely amounts to overcompensation of the affected industries. These papers use various simulation methodologies to forecast potential impacts of carbon taxes or caps. Bovenberg and Goulder (2002) and Goulder, Hafstead and Dworsky (2010) utilize general equilibrium models to assess the likely impacts of a carbon tax and various cap-and trade policies on a wide set of industries. Burtraw and Palmer (2008) simulate the US electricity sector under potential cap-and-trade scenarios. Smale, Hartley, Hepburn, Ward and Grubb (2006) simulate several industries under a carbon cap in Europe using an assumption of Cournot competition. All these studies find that for many industries, compensation of less than 20 percent of emissions would offset the profitability impacts of regulation.

In this paper we study impacts on firms of the largest, in monetary terms, cap-and-trade market in the world - the European Union's Emissions Trading System (ETS) for CO_2. This is, to date, the most significant effort by far at regulating CO_2 emissions in the world. As a role model for carbon cap-and-trade, the ETS has been closely scrutinized both within and outside the European Union. From the outset, the relative impact of the ETS on EU industries has been a controversial topic, one that has strongly influenced policies for the allocation of emissions allowances. During its first phase of operation from 2005 through 2007, the prices of emissions allowances in the EU market were quite volatile. While this volatility has sparked criticism about the design and implementation of this phase of the

[2]For example, Ryan (2007) demonstrates how the Clean Air Act significantly increased the sunk cost of entry in the Portland cement industry. Puller (2006) demonstrates how firms can profit from increased regulation by raising rival's costs, leading them to promote the adoption of those regulations.

market, we take advantage of it in order to examine the impact of CO_2 prices on firms.

Rather than attempting to directly untangle the many competing effects of the ETS on firms, we focus on the stock market valuations of public-traded firms subject to CO_2 regulation. Specifically, we examine the impact of a sharp devaluation in CO_2 prices in late April 2006 as an event study on the share prices of affected firms. Such an exercise can be interpreted in several ways. Under an assumption of fundamental market valuation these prices should reflect the market's expected discounted future profits of the firms. Even if one does not adhere to an assumption that the market fully reflects expectations of future profitability, the event provides a useful window into the beliefs of the market about the impacts of movements in CO_2 prices.

Our results imply that rather than being hurt by the imposition of CO_2 regulation, several industrial sectors benefited from the ETS. Indeed the sharpest declines in equity prices occur within industries that are the most carbon intensive, or electricity intensive. Such a response indicates that CO_2 prices play a significant role in determining product prices and revenues in many of these industries. We also examine the responses in relation to a measure of international trade exposure, and find weak evidence that the benefits of higher CO_2 prices were concentrated amongst sectors with little exposure to international trade.

In section 2.2, we develop a simple model of the impacts of CO_2 costs on firm profitability in order to illustrate the potential impacts. In section 2.3, we briefly review the EU CO_2 market and its pricing from 2005-07 and examine the impact of the crash in permit prices in late April 2006. In section 2.4, we examine the underlying elements of firm characteristics that influenced the response to the change in CO_2 prices. We conclude in section 2.5.

2.2 Emissions Regulations and Firm Profits

In this section we develop a theoretical model considering the potential impacts of environmental regulation, or more specifically emissions costs, on firm profitability and performance. The model provides a useful framework for decomposing and illustrating the various potential impacts, both positive and negative, of emissions costs on firms. Consider a firm producing products for a market represented by the demand curve, $P(Q)$, where Q represents total industry production in this market. The firm is subject to cap-and-trade regulation of its emissions, which are in turn a function of its emissions rate, r, and its total production, q. We assume that the production technology determines the emissions rate, $r(q)$ and that this rate cannot be changed over the time horizon we are considering. The per-unit price of emissions allowances is τ, resulting in direct compliance costs of $\tau r(q)q$. However, the firm may posses allowances A equal to its initial allocation less net sales. Considering both input and environmental costs, the profits of this firm, i, can be represented as:

$$\pi_i = P(Q)q_i - C_i(q_i, \omega) + \tau A_i - \tau r_i(q_i)q_i$$

where the function $C_i(q_i, \omega)$ represents the total cost of producing q with a vector of input costs, w. The impact on profits of a change in the allowance price, τ, can be expressed as

$$\frac{d\pi_i}{d\tau} = P\frac{dq_i}{d\tau} + P'\frac{dQ}{d\tau}q_i - \frac{\partial C}{\partial q_i}\frac{dq_i}{d\tau} - \frac{\partial C}{\partial \omega}\frac{d\omega}{d\tau} + A - rq_i - (r'q_i + r)\frac{dq_i}{d\tau}\tau. \tag{2.1}$$

For firms with market power in their product market, we can also consider the effect on product prices to be a combination of changes in their own output and the output of other firms. For firm i

$$P'\frac{dQ}{d\tau} = P' * \left(\frac{dq_{\neq i}}{d\tau} + \frac{dq_i}{d\tau}\right)$$

where $q_{\neq i}$ is the output of all other firms, $q_{\neq i} = Q - q_i$.

Assume firms maximize profits with respect to q. Define $\pi_i^* = \pi_i(q_i^*, q_{\neq i})$. For shocks that have marginal influence on q_i, the envelope theorem implies,

$$\frac{\partial \pi_i^*}{\partial q_i} = P + P'q_i^* - \frac{\partial C}{\partial q} - (r'q_i^* + r)\,\tau = 0. \tag{2.2}$$

In other words, the change in profitability though own output would be negligible. However, there are still effects relating to direct costs, the value of allowance holdings, and changes in market prices due to the responses of other firms in the industry. Combining equation (2.2) and equation (2.1),

$$\frac{d\pi_i^*}{d\tau} = P'\frac{dq_{\neq i}}{d\tau}q_i^* + -\frac{\partial C}{\partial \omega}\frac{d\omega}{d\tau} + A - rq_i^*. \tag{2.3}$$

The individual terms in equation (2.3) illustrate the competing potential effects of a change in the allowance price. First, revenues may increase due to the fact that other firms in the industry have collectively responded by reducing output. This is similar to a "raising rivals' costs" effect.[3] Under the assumption that firms would reduce output in the face of an increase in allowance costs, this term would be positive. Second, the middle term on the right hand side of (2.3) captures the impact of changes in input costs due to a change in the allowance price. To the extent that these inputs (*e.g.* electricity) come from industries that are themselves subject to the environmental regulation, this term would presumably be negative. The last term, $A - rq$ reflects the change in direct compliance costs of a change in allowance prices. If a firm is "short" in allowances, then $A < rq$ and this term would be negative.

The model is intended to be general, encompassing both perfectly competitive industries and those in which individual firms have market power. However, it is important to also acknowledge aspects of oligopoly competition that are not explicitly represented within this framework. In oligopoly settings, cost shocks such as environmental regulations can increase profitability by increasing the severity of market power in an industry. In a dynamic setting,

[3]Salop and Scheffman (1983)

the environmental regulation could serve as a barrier to entry or even as a collusive focal point. Even in a static setting, the imposition of an environmental tax can increase margins under certain demand structures (Seade 1985).

In the following sections, we will examine each of these potential effects empirically. The relative magnitudes of these effects will largely depend upon three key factors, the elasticity of demand for the firm's product, the firm's endowment of permits, and the relationship between a firm's marginal cost and its average cost with respect to emissions and other input prices. Figure 2.1 helps to illustrate these factors. We assume here that a firm faces a residual demand curve D, and has a marginal cost function $c_{\tau 1}$ before the imposition, or increase, in allowance prices. In this figure, we also assume that the residual demand curve D for this firm is unaffected by a change in allowance prices, one condition for which is that all of the firm's competitors operate outside of the capped region.

The classic analysis of the incidence of taxation on such a firm would imply a vertical shift of the marginal cost curve to $c_{\tau 2}$. In the context of environmental regulation, this is equivalent to assuming that emissions rates are constant for all production quantities. If true the producer surplus is clearly reduced from the sum of areas B and C to the area A in Figure 2.1a. The allocation of revenues collected, or of permits, would then be critical in determining the net effect of the regulation. If the firm received a free allocation equivalent to 100% of its ex-post emissions, this would be a transfer equivalent to the areas C and D, which totally offsets the increased regulatory cost. As long as the demand for product is sufficiently inelastic, the firm's net profit improves because its revenue increases without any increase in environmental costs. Indeed as Bovenberg and Goulder (2002) demonstrate, only a relatively small allocation of emissions allowances is necessary to fully compensate many industries for changes in profits due to CO_2 costs.

However, even without an allocation of allowances, the impact on firm profits can be ambiguous. This is due to the fact that there are both heterogeneous firms and production technologies within most industries. Consider a case where emissions rates are increasing with production quantities, as illustrated in Figure 2.1b. The increase in allowance costs now raises marginal costs, and therefore prices in this perfectly competitive circumstance. The increase in *average* costs is well below the increase in marginal costs, however. Now the new producer surplus, area A, could be larger than the previous surplus of B and C. A similar, even larger, effect could arise if an individual firm happens to have a "cleaner" technology than its rivals. Such a circumstance would have the effect of decreasing the residual elasticity of demand for the clean firm. Again product prices could rise much faster than average production costs.

Of course, such an effect strongly depends upon the fact that much of the incidence of increased emissions costs are being passed on to consumers. If the firm in question were instead faced with very elastic demand for its product, even a substantial convexity in the marginal cost curve could not compensate for the fact that the producer is absorbing the bulk of the emissions cost increase (Figure 2.1c).

This discussion is meant to illustrate the varied potential effects and emphasize the importance of several key industry characteristics in determining the net effects of environmental regulations. In the following section, we develop several proxy variables meant to reflect these characteristics in order to examine the market return of individual firms and industries in response to a substantial decline in emissions costs.

2.3 The EU Emissions Trading System

The EU Emissions Trading System (ETS) was developed as one of the central mechanisms for which the European Union member states could achieve compliance with the commitments under the Kyoto treaty and is in many ways a remarkable accomplishment. The world's first significant cap-and-trade system for CO_2, the ETS covers over a dozen industries and 27 countries, including several that took on no Kyoto obligations. The ETS has been rolled out in phases. The first phase, running from 2005 through 2007, was intended as much to develop institutions and gain regulatory experience as to achieve substantial CO_2 reductions. The overall cap for the market was an aggregation of caps developed by each participating country through their "national allocation plans," previously analyzed by Betz, Eichhammer and Schleich (2004). The EU established guidelines for the development of these plans, but member states were left with significant latitude. Efforts at setting an appropriate cap were complicated by the fact that, prior to 2005, the monitoring of CO_2 emissions of many facilities and countries was unreliable at best. Caps were supposed to be set in a manner that would place emissions reductions on a trajectory consistent with meeting the Kyoto targets. However, the effective stringency of the Kyoto targets varies greatly amongst EU member states, and the implementation plans themselves reflected large differences in these goals, as well as in the relative weight countries chose to give to the capped sectors covered by the ETS as opposed to those sectors counted under Kyoto but not under the ETS.

A second source of diversity amongst participating nations was their relative approach to assigning permits to the covered sectors. As chronicled in Ellerman and Buchner (2007), Kettner, Koppl, Schleicher and Thenius (2008), and Ellerman, Joskow and on Global Climate Change (2008), countries such as Spain, Italy, and the UK appear to have imposed more stringent caps and as a consequence the affected industries in these countries, particularly in the power sector, were allocated few permits than their observed emissions. These firms were therefore net buyers of permits within the EU. Industries in other countries, particularly in Eastern Europe, were observed to emit far less than their allocations.

Another important contrast lay in the allocation of permits across the various industrial sectors. Although there were differences in countries' approaches to the allocation of permits to their industries, some common themes emerge. In general, many regulated firms in the manufacturing sectors received more permits than they subsequently needed to cover their observed emissions. Those providing power and heat, most notably electricity firms, were generally "short" of permits, but still received allocations equivalent to a substantial majority of their emissions.

Overall, by the end of phase I, available permits exceeded measured emissions by about 2.8%. Although the eventual surplus in permits led to a perception of intentionally lax regulation through "over-allocation," the picture is more nuanced. An ex-post realization of a surplus does not necessarily imply over-allocation, since a surplus of allowances can arise from either over-allocation or over-abatement. Since emissions prices were quite high for some of this period, it is natural to expect some abatement to have occurred, at least while emission prices were high. Studies by Ellerman and Buchner (2007) as well as Delarue, Voorspools and Dhaeseleer (2008) indicate that at least some abatement did take place. In addition, macroeconomic and weather shocks may have played a role in lower than expected emissions, and specific directed regulations such as aggressive subsidies for renewable electricity production may have been sufficient to tip the market into surplus.[4] Importantly, none of this was known for much of the first phase, and it was only after the phase was more than 2/3 complete that the surplus conditions pushed emissions prices to near zero.

2.3.1 ETS Market Performance

The most notorious aspect of the ETS during phase I was the volatility of the permit prices, which was greatly exacerbated by the fact that permits could not be "banked" for use beyond 2007. The ETS market was characterized by an early period in which prices were higher than anticipated and a later period in which the price eventually reached zero in the face of a surplus of permits that held no value beyond 2007. From the onset of trading in January through March 2006, prices rose steadily to over 30 Euro/ton. While this price rise appears somewhat surprising in hindsight, given the eventual surplus of permits, it was not necessarily considered anomalous at the time. Many attribute the relatively high prices during this phase to the fact that prices for natural gas, which largely defines the marginal costs of reducing CO_2 emissions in the power sector through its substitution for coal, were steadily rising during this period.[5] In addition, while firms from countries "short" on permits were apparently relatively active in trading from the beginning, those from many "long" eastern European countries were not due to delays in integrating the regulatory platforms with that of the EU. This may have contributed to masking what later emerged to be a surplus of available permits.

The lack of reliable information about aggregate emissions was also a critical contributor to the uncertainty about price levels. This changed on April 25, 2006 when the first reports of country level emissions began to leak into the permit market. As can be seen in Figure 2.2, the reaction was dramatic. Over the next few days, the permit price as reported on the European Climate Exchange fell from 28 Euros/ton on April 25 to 14 Euros/ton on April 28. The price drop hit both phase I permit prices as well as permits covering phase II, which

[4] See Convery, Ellerman and De Perthuis (2008)

[5] Ellerman et al. (2008)

had begun trading in 2006. In fact, the surpluses reported during those periods were not reflective of the more modest surplus left at the end of phase I, and even these initial reports were revised shortly after they were made public. By May 15, when the final emissions totals were officially released, phase I prices had rebounded and then fallen slightly again to settle around 16 Euros/ton.

During this one month period, the general movements of prices for both the phase I and phase II permits had been generally consistent with each other, although the magnitudes were more muted in the case of the longer-term phase II permits. Later in 2006 the two prices series diverged for good, with the phase I prices starting a steady decline toward zero and the phase II series settling into a range around 20 Euros/ton.

2.3.2 Equity Market Effects

We now turn to the question of how the sharp devaluation in permit prices in April 2006 impacted expectations about firm profitability. A few papers have empirically looked at different segments of the EU market. Sijm, Neuhoff and Chen (2006) examine the implications specifically for electricity prices in the Netherlands and Germany and find substantial pass-through of carbon cost. Convery et al. (2008) note that net incomes of several large electricity producers increased throughout phase I of the ETS. Two similar papers, Veith, Werner and Zimmermann (2009) and Oberndorfer (2009) examine stock market returns of electricity companies using a panel regression of share prices on CO_2 prices throughout the phase I period. Both find that share prices of large electricity producers who were regulated under the ETS were positively linked with prices for CO_2. However, Veith et al. (2009) find that share prices of "clean" electricity producers not covered under the ETS had no significant response to CO_2 prices.

In this paper we also utilize equity prices of publicly traded firms. It is important to note that many firms directly subject to the CO_2 cap, as well as those in impacted industries, are privately held or government owned. A large number of publicly traded firms were also effected, however, and we focus our attention on these firms. We employ a standard event-study approach.[6] We examine firms contained in the Dow Jones STOXX 600 index, which is similar to the S&P 500 but covers European firms.[7] We focus on the three days after the initial leak of permit market information, the daily returns for April 26-28. Several papers have utilized an event study approach to assess the impact of environmental regulation on firm profits, including Kahn and Knittel (2003), Linn (2010), and Linn (2006). Because this approach has usually utilized a political or legal decision as the "event," a common concern has been that information may have leaked into the market before the examined event date.

[6] Fama, Fisher, Jensen and Roll (1969); more recent surveys include Brown and Warner (1985) and MacKinlay (1997)

[7] We chose this index because of its breadth of firms and of geography. Other commonly cited European Indices such as the FTSE 100 and the DAX are more limited in coverage of European countries and industries.

Here we can be confident here that there was little leakage of information as this information would have impacted the CO_2 price, which was steadily rising up until our event date.

We utilize the following specification for investigating the potential for extraordinary returns during this event window.

$$\ln(S_{i,t}/S_{i,t-1}) = \alpha_i + \beta_i \ln(M_t/M_{t-1}) + \gamma_i EVENT_t + \epsilon_{i,t} \tag{2.4}$$

where $S_{i,t}$ is the share price of firm i and M_t is the price of the market index at time t, and $EVENT_t$ is a dummy variable that is scaled according to the length of the event window. For our base specification, where the event window is 3 days, $EVENT_t$ will be scaled by 1/3 so that $gamma_i$ represents the cumulative excess return during the event window.

We run regression (2.4) for each stock in the index individually, and aggregate individual γ_i to summarize results by industry categories. We perform this aggregation through the following regression.

$$\widehat{\gamma}_i = \theta_j + \varepsilon_i \quad \forall i \in j. \tag{2.5}$$

Industrial categories j are based upon NAICS 2 digit classifications.[8] Intuitively, the coefficient value θ_j therefore represents the average effect of all firm specific impacts within each industry sector.

Table 2.1 summarizes the event effects by industrial classification. Many of the largest significant declines were registered in industries that feature prominently in the EU ETS, including Mining and Oil & Gas Extraction and Utilities. However, there are also notable declines in such industries as Real Estate, Accommodation & Food Services, and Construction. As we describe below, each of these industries are relatively large users of electricity and sell to relatively local markets. The largest increase was in Wholesale Trade.

These results are merely meant to summarize general effects. The groupings in Table 2.1 are somewhat problematic, as classifications can be imperfect and there can be considerable heterogeneity of firms within a classification. This latter fact is highlighted by Table 2.2, which summarizes the effects for firms contained in the Electricity sector, using auxiliary data on electricity generation units from the Carbon Monitoring for Action project (carma.org) published by the Center for Global Development, Washington DC.

The second column of Table 2.2 presents the event coefficient for each firm, while columns 3-5 summarize some key characteristics of the firms. When one bores down into the detailed characteristics of a firm, as is more easily done within the electricity sector, some suggestive patterns begin to emerge. In general, the biggest declines were concentrated within firms who produce electricity with relatively low CO_2 emissions, such as the hydro or nuclear intensive firms Fortum, British Energy, and Electricite de France. Some coal intensive firms

[8] These data are provided by Compustat. Thompsons Datastream provides a classification called INDM which provides similar results as the 2-digit NAICS, but NAICS was chosen because it is more widely used in the literature and because it is more easily linked to other industrial characteristics discussed below. However, Weiner (2005) evaluates several industrial classification schemes and finds drawbacks in each.

such as Drax and RWE registered declines, but they were more modest than those of the "clean" producers. Last network operators such as National Grid and Red Electrica, with no position in the production or sale of electricity, registered almost no impact.

These results are consistent with an explanation of the effects that emphasizes the importance of revenue impacts in the product markets. All the firms in Table 2.2 who sell bulk electricity experienced declines in revenues, and only some experienced significant declines in production costs. Many of these firms were also substantial holders of emissions permits at the time of the crash in permit prices. In the following section we develop several more general indices meant to capture the relative sector level and firm specific characteristics that could influence the permit price effects and test their relevance on market returns during this event period.

2.4 Testing Determinants of Profitability

In this section, we examine which industry and firm characteristics determine the profitability of some firms in the face of CO_2 price changes. First we test the importance of firms' allocation of permits, net of emissions, in determining abnormal returns. Then we test whether the share price changes described in the previous section are consistent with a "revenue effect."

2.4.1 Asset Value of Permit Holdings

We first examine the effect of permit allocation, and emissions on the performance of share prices during the event. For this task we utilize the emissions data contained in the EU's Community Independent Transaction Log (CITL). This dataset contains facility level information on the allocation and emissions of over 12,000 facilities throughout the EU. Unfortunately, firm ownership of facilities is reported inconsistently within the CITL, making necessary a manual matching of facilities to firms, and then to individual stock listings.

We were able to match 90 publicly-traded firms in the largest sectors regulated by the ETS. For each of these firms, we take total 2005 emissions and permit allocations aggregated over all covered facilities owned by the firms.

We examine whether these firms' permit allocations and emissions explain abnormal returns. Given a drop in permit prices, those firms with positive net permit positions will lose more profits than others with a negative net position, all else equal. In theory this will be reflected in the stock price. We test this by estimating the following equation:

$$\widehat{\gamma}_{ij} = \theta_j + \mu(A_i - E_i)/M_i + \eta_{ij}, \tag{2.6}$$

where A_i be the historic 2005 allocation, E_i be historic 2005 emissions (as measured in the spring of 2006), and M_i is the firm's historic market cap in Euros (*e.g.*, on April 25, 2006). In order to control for industry average differences, we examine including industry fixed effects.

Note that, although the CITL registers all transactions, only the allocations and emissions data are currently publicly available. Therefore we do not know the actual holdings of a given firm on any day, only their initial allocations. Our values for $(A - E)$ should be considered only proxies for the "true" net position of firms at the time of the event. Importantly, the broader market also did not know these "true" net position, and was relying upon the same data, which were finalized on May 15, that we utilize here.

The net permit position $(A_i - E_i)$ is normalized by market capitalization. This is done because larger firms could have greater variation of net permits. Furthermore, this normalization implies a μ coefficient of the change in market capitalization given a change in net permits.

If profit impacts were driven completely by net emissions costs, we hypothesize that the coefficient μ would equal roughly the drop in permit price times three, or about -42. A firm with, say, 1 million tonnes of excess permits in 2005 may be expected to have extra permits in 2006 and 2007. The value of these unused permits fell by the drop in the permit price, which was around 15 Euros per tonne. Hence, this hypothetical firm would have lost, 1 million tons/year * 3 years * -14 Euro, or 42 million Euros.

Table 2.3 reports the results. We find that the coefficient on net position is statistically significantly different from -42. In fact, we do not even find a statistically significant coefficient. In Panel A, we exclude fixed effects and find a coefficient of -6.9 that is insignificant. Even after controlling for industry fixed effects, in Panels B and C, we find a very similar result (negative and insignificant).

Given the lack of market information about permit trading, investors were unlikely to know the exact net position of firms, and may have had difficulty even estimating the sign of net position. Figure 3, which plots the 90 firms' permit allocation and emissions during 2005, demonstrates this point. Many firms had been allocated permits that were very highly correlated with their 2005 emissions levels. We find that initial allocation explains over 95 percent of the variation in 2005 emissions.

In Table 2.3, we next examine whether the abnormal returns were correlated with a firm's level of emissions, or allocations. We find no evidence of this in Panel A. However, the picture becomes more clear once we control for industry fixed effects. As described above, many industry classifications were "long" in permits during this period. The important exception is the power industry which was on net short of permits. We therefore estimate the power industry, as the one segment known to be short, separately in Panel B. In Panel C, we estimate the influence firm-level emissions and allocations on all other industries, controlling for industry fixed effects.

With industry fixed effects a clear distinction between the power sector and other industries emerges. Within the power sector, firms with high levels of emissions outperformed the "cleaner" firms when the allowances prices fell. There is a strong relationship between emissions and changes in market capitalization, with each ton of emissions improving market cap by 6.25 Euros. Firms with higher allocations also had better returns, but recall that emissions and allocations are almost completely co-linear, so this is likely also an emissions effect. Firms in the other industrial sectors, which were net long on permits, experienced the opposite effect. Firms with higher allocations suffered the largest declines when the permit price fell, with each added ton of allocation implying a reduction of 31.5 Euros in market capitalization. As with the power sector, both emissions and allowances produce nearly identical coefficients, reflecting the strong correlation of these two variables.

This firm-level analysis of permit holdings and emissions implies that, within industries that were net long on permits, dirtier firms suffered the largest declines. This is consistent with a market expectation that these firms had suffered the largest decrease in aggregate permit asset value, as these firms were the largest holders of permits within their industries, and their asset values in permits exceeded their emissions liabilities. For the power sector, it is the cleanest firms that suffer the most. This is consistent with a market focus on the impact of permit values on electricity prices, combined with a view that dirtier firms experienced a net decline in their abatement costs to somewhat offset the decline in product prices. These dirty firms in the power sector still experienced abnormal negative returns, but they were more modest declines than those of the cleaner firms.

2.4.2 Tests of Revenue Effects

Recall from Section 2.2 that the revenue effect depends on how a cost shock in an industry affects the output prices, $\partial p/\partial \tau$. This in turn will depend upon the elasticity of demand for the product, the convexity of a firm's costs with respect to emissions costs, and the relative emissions of other firms in the industry. For example, industries that have little international trade exposure, use many dirty inputs, and produce substantial carbon emissions are more likely to have a strong revenue effect. In order to test the importance of these factors, we examine the abnormal returns during the event window as estimated in equation 2.4, $\widehat{\gamma}_i$.

$$\begin{aligned} \widehat{\gamma}_{ij} &= \delta_0 + \delta_1 1(DO_j > 0) + \delta_2 DO_j + \delta_3 1(DI_j > 0) \\ &\quad + \delta_4 DI_j + \delta_5 1(TE_j > 0) + \delta_6 TE_j + \nu_{ij}, \end{aligned} \tag{2.7}$$

where DO_j is a measure of how dirty (carbon intensive) is an industry's output, DI_j is a measure of how dirty are an industry's inputs, and TE_j measures the trade exposure of the industry. We describe each of these variables in more detail below.

Sectors were characterized by the "dirty output," "dirty input," and "trade exposure" variables at the NAICS 3-digit level. Dirty output (DO) comes from combining CITL emissions data with Thomson's Datastream financial data. For all sectors j where at least one firm was matched in the CITL, DO is given by the following formula.

$$DO_j = \frac{\sum_{i\in(j\cap CITL)} Emit_i}{\sum_{firm\in(j\cap CITL)} S_i} \tag{2.8}$$

where $Emit_i$ is the sum of facility level emissions in the CITL over all facilities owned by firm i and S_i is the 2005 revenue of firm i. The subscript j indexes NAICS3 sectors, and $CITL$ indexes firms contained in the CITL emissions data set. The emissions factor calculated above is then normalized to the 0-1 range. Emissions intensity for *any* firm in a given NAICS sector will therefore be based upon the measured emissions of firms matched with CITL data in that NAICS sector. There were 90 firms for whom we have been able to match with the facility level emissions data, and 202 firms contained in the STOXX 600 index drawn from the sectors for which we have matched emissions data.

Dirty input comes from input-output tables of industrial activity. DI_j is the direct plus indirect input use of the electricity sector in producing one dollar of output in sector N. We are not aware of sources of input output tables for the EU with NAICS nomenclature, so the index here is calculated using US figures from the Bureau of Economic Analysis (BEA). As with DO, we normalize the value of DI to range between 0 and 1.

Trade Exposure (TE) is a measure of how much a given commodity is internationally traded. We use a measure of Trade Exposure that the European Union has proposed to be used in determining which sectors get free allocation due to industrial competitiveness concerns.[9]

$$TE_j = \frac{(EXPORT_j + IMPORTS_j)}{(OUTPUT_j + IMPORTS_j)}$$

$EXPORTS$ and $IMPORTS$ are with respect to the EU region, so intra-EU trade (which is uniformly under the ETS) is not counted. US trade (from COMTRADE) and production (from BEA) data was used to construct these measures. Though European data is preferable, US data should be equivalent if US and European input-output tables and trade profiles are similar. US data were used because they were already coded to NAICS, whereas European data are categorized by NACE codes, which require further (imperfect) translation to NAICS via correspondence tables.

Table 2.4 provides the summary statistics for twenty sectors (based on two-digit NAICS codes). For each sector, the table reports average abnormal returns during the event window. In addition, the sectors' industry characteristics (DO_j, DI_j, and TE_j) as well as the market capitalization are summarized. The mining, oil and natural gas extraction sector is that which is most electricity intensive: it had the largest average abnormal stock drop of approximately 2.7 percent. Utilities have the highest carbon emissions intensity: its average stocks had an abnormal decline of about 1.8 percent.

[9]Convery, *et al.*, 2008

Note that for each of these variables there are many sectors with no value. For DO this is because many industries are not covered under the cap and trade system. In the sample of 600 firms, roughly 40% are in industries covered by the ETS and therefore have non-zero values for DO. In the case of DI, there are some (roughly five percent) firms with NAICS codes not contained in the BEA input-output tables. In the case of trade exposure, this is an artifact of our reliance on trade and production data. These data are focused on the manufacturing sectors, and therefore several industries, particularly service oriented ones, are not considered to be involved in international trade. About 60% of the 600 firms have no value for trade exposure. It is because of these issues that we include dummy variables that are applied to all firms with non-zero values for DO, DI, and TE respectively in the specification described above.

Table 2.5 reports the results of different variations of regression 2.7. The first two columns report the results controlling only for dirty output, or dirty input respectively. The third column controls only for trade exposure. The fourth and fifth columns interact DI and DO with trade exposure, under the intuition that trade exposure should matter less in relatively "clean" industries that are unaffected by CO_2 prices. Column seven combines all these variables by interacting both DO and DI with trade exposure.

From Table 2.5, it is clear there is a relationship between carbon intensity and performance during the event window. Firms from industries with high emissions (large DO) or relatively dirty inputs (*e.g.*, high electricity usage) saw their share prices decline. This is suggestive of a revenue effect, as firms in these industries will have experienced a decline in their competitor's, as well as their own, marginal costs. When DI is interacted with trade exposure, the coefficient on DI roughly doubles, suggesting that it was firms with no trade exposure who are largely driving the negative value on DI. The interaction term on DI and TE is positive, but very imprecisely estimated. When all terms are considered simultaneously, higher values of both DO and DI significantly impact a decline in share prices during the event.

It might at first seem counter-intuitive that the firms most directly impacted by CO_2 regulations would be the greatest losers from a decline in CO_2 prices. Keep in mind that these values are measuring the relative carbon intensities of *industries*, not the individual firms within industries. Thus we interpret these results as being consistent with the hypothesis that product prices, and therefore revenues, were negatively impacted by the CO_2 price shock. Although costs were also reduced, either through the direct or indirect exposure to CO_2 regulation, it appears that the revenue effects were stronger. For regulated industries, this is almost certainly a consequence of the fact that allocations were closely linked to emissions, as illustrated above. For these firms, the revenue effects would naturally be the strongest as the reductions in costs are largely offset by a concurrent reduction in the value of permit holdings.

We examine the robustness of these results in several ways. One question is the appropriate time window for the event. This is particularly true as the volatility in permit prices continued beyond the 3 day window examined above. To address this question we also exam a 30 day event window we call $BIGEVENT$, consisting of 5 days prior and 25 days after April 25, 2006. We generate new $\widehat{\gamma_i}$ estimates using the $BIGEVENT$ window and perform the same analysis on the influences on share price performance. Table 2.6 describes the results for these regressions. As before, both DI and DO produce negative, although insignificant, coefficients when considered on their own. When all factors are included (column 7), the coefficients for dirty inputs and dirty outputs are negative and significant at the 10% level. Interestingly, the impacts of trade exposure are much stronger than during the shorter event window. While firms with trade exposure in general saw a decline in shares, the interaction terms for both DO and DI are positive and at least weakly significant. This indicates that although dirty firms saw a decline in shares overall, the dirtiest firms that were most exposed to international trade benefitted from the CO_2 price decline.

In Table 2.7, we add a measure of the firm's debt-to-equity ratio. Note that the net present value of all future profits equal the sum of equity and debt. By including the debt-equity ratio, we test the robustness of our results that the findings are representative of changes in profits, not just equity. Although debt-to-equity is a significant factor, it does not change the underlying picture with regards to dirty inputs and outputs during the short event window. In Table 2.8, we test the importance of the CAPM framework to the results by testing the event on the unadjusted returns (e.g. no β term) of the shares. The results are very similar to those of table 5.

In Table 2.9, we test for the presence of possible spillovers to a neighboring market by performing a similar analysis for the stocks in the US Standard and Poors 500 index. When all factors are considered, the only variable with a significant impact on returns is the DO index variable, which is positive, indicating that dirty firms experienced an increase during this period. One possible interpretation is that the event in the EU lowered expectations about the probability or the cost of future regulation in the US. In Table 2.10, we analyze a similar time frame from the year 2004, a date *before* the EU CO_2 market came into existence, as a form of falsification test. Although certain characteristics were significant in determining the abnormal returns of shares during this 2004 period, the results are quite different from the results from the 2006 CO_2 price crash.

2.5 Conclusions

The development and application of any significant new environmental regulation will involve some level of debate over its economic impacts. This is particularly true in the case of regulations to combat climate change because the stakes are so high. The annual value of permits consumed in the European ETS market we study reached nearly \$60 Billion. A market in the United States would be 2 to 3 times the size of the European market. These values are an order of magnitude larger than any other previous emissions trading markets. These sums have generated intense interest in the potential incidence of these costs, and many industries are making the case for some form of free permit allocation to offset these costs.

However, the cost impact is only one part of the story from the perspective of firms and industries. The impact of emissions costs on revenues is another critical consideration. It a desire to examine this full portfolio of impacts that has drawn us to examine the European ETS market. We have used an event-study approach to analyze the response of the stock market to the devaluation of CO_2 permit prices in late April 2006. This provides one of the first opportunities to empirically test the impacts of CO_2 regulation on major industries and firms. By looking at the impact of a sharp decline in CO_2 prices on the equity prices of impacted firms, we can get a strong sense of what the market believes to be the net impacts of CO_2 regulations.

The story that emerges from an examination of this event is that the equity markets were strongly focused on revenue effects. Our results demonstrate, fairly robustly, that the share prices of firms from the "dirtiest" industries experienced the largest abnormal declines during this period. For firms that are directly regulated under the ETS program, consideration of permit holdings almost certainly influenced investor response. Although our data on allocations appear insufficient to explicitly identify a "net holdings" effect, we do find evidence that allocations played a role in the market's response to the CO_2 price crash.

Within the power sector, which was as a whole "short" of permits, the share prices of firms with the highest emissions rates, perform better than the "cleaner" firms within this sector. The share prices of many of these high emissions firms did experience abnormal declines, but these declines were less severe than those of their low carbon intensity competitors. The fact that very low-carbon emissions firms declined the most gives strong indication of the market's focus on how declining CO_2 prices would reduce the revenues of these firms through lower electricity prices. The fact that the high emissions firms still experienced declines highlights the fact that the market also understood that these firms were holding large portfolios of allowances and experienced a loss in that portfolio that largely offset their cost savings from lower CO_2 prices. Within other industries that were in aggregate allocated more allowances than were consumed, those firms with the largest allowances experienced the largest abnormal declines.

It is important to recognize the many caveats that must be applied to interpreting these results. The ETS was a very new market, which was one of the causes of the volatility we utilize here. It would be heroic to assume that the stock market completely and accurately processed the information that emerged in late April 2006. In addition, while the crash affected both near-term and long-term CO_2 prices, the impact on the near-term Phase I prices was much more pronounced. The events of 2006 may also have impacted expectations about future allocations of emissions permits, as well as expectations about prices. Because our event study uses the same time window for all stocks, any contemporaneous events could also be causing the abnormal returns. We looked for sector-specific announcements in this period. Specifically, oil prices did not change dramatically.

Nonetheless, these results are largely consistent with what simulation studies had predicted could be the case for many of these industries. These studies forecast an increase in revenues that would largely offset the increase in regulatory costs. In fact, our results imply that for clean firms in dirty industries, these revenue effects are larger than cost increases. These are important facts to bear in mind when setting policies regarding allocations to impacted industries. In many cases, those directly or even indirectly impacted by CO_2 costs may need little compensation. Instead it is their customers who will be most affected.

2.6 Figures and Tables

Figure 2.1: Theoretical Change in Producer Surplus under Environmental Regulation. Under a tax, or auctioned permits, firms gain area A but lose areas B and C. However, if firms are allocated permits equal to their equilibrium emissions, they gain A and D and lose only B.

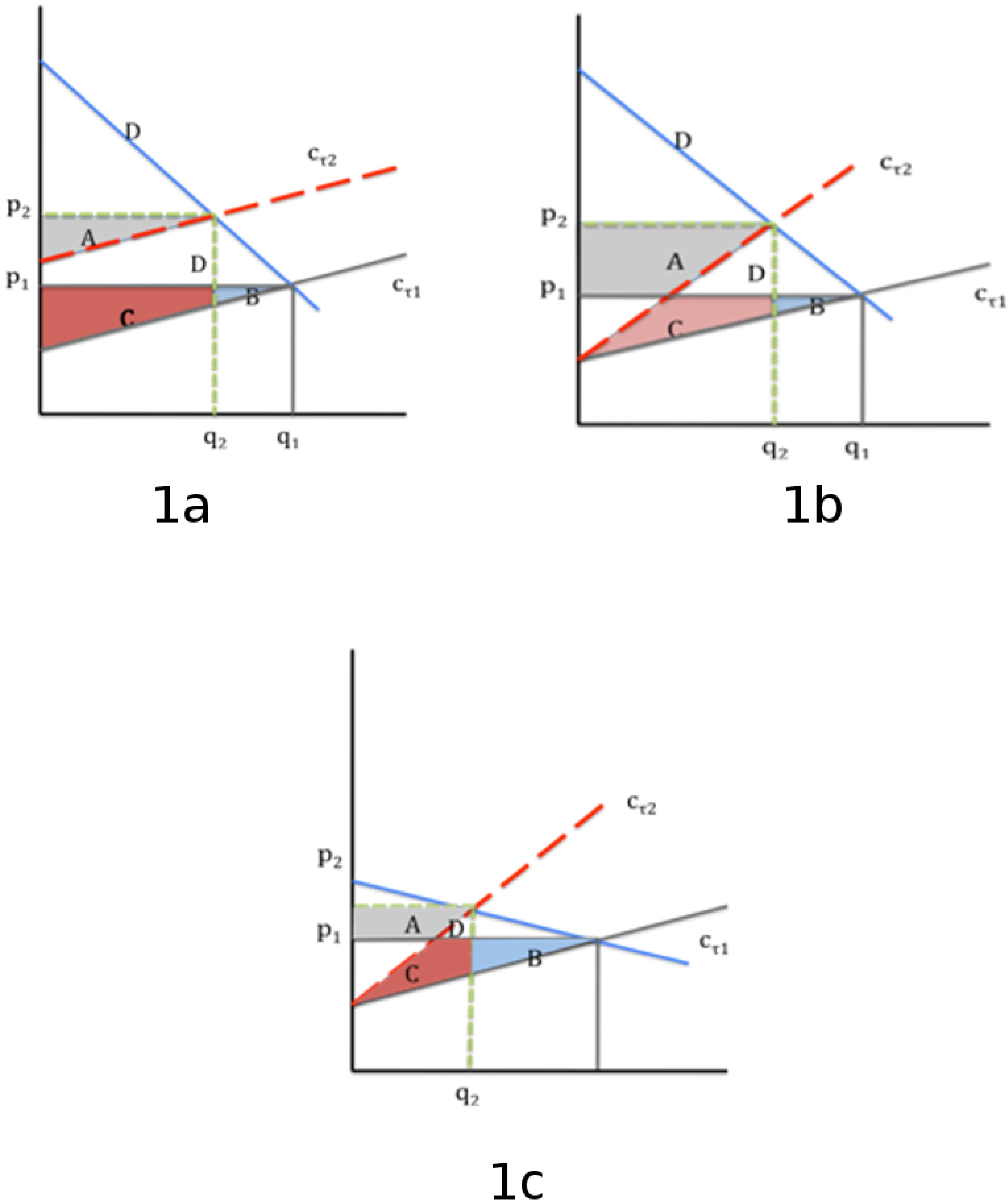


Figure 2.2: EU Carbon Prices, Stock Index, and Oil Prices

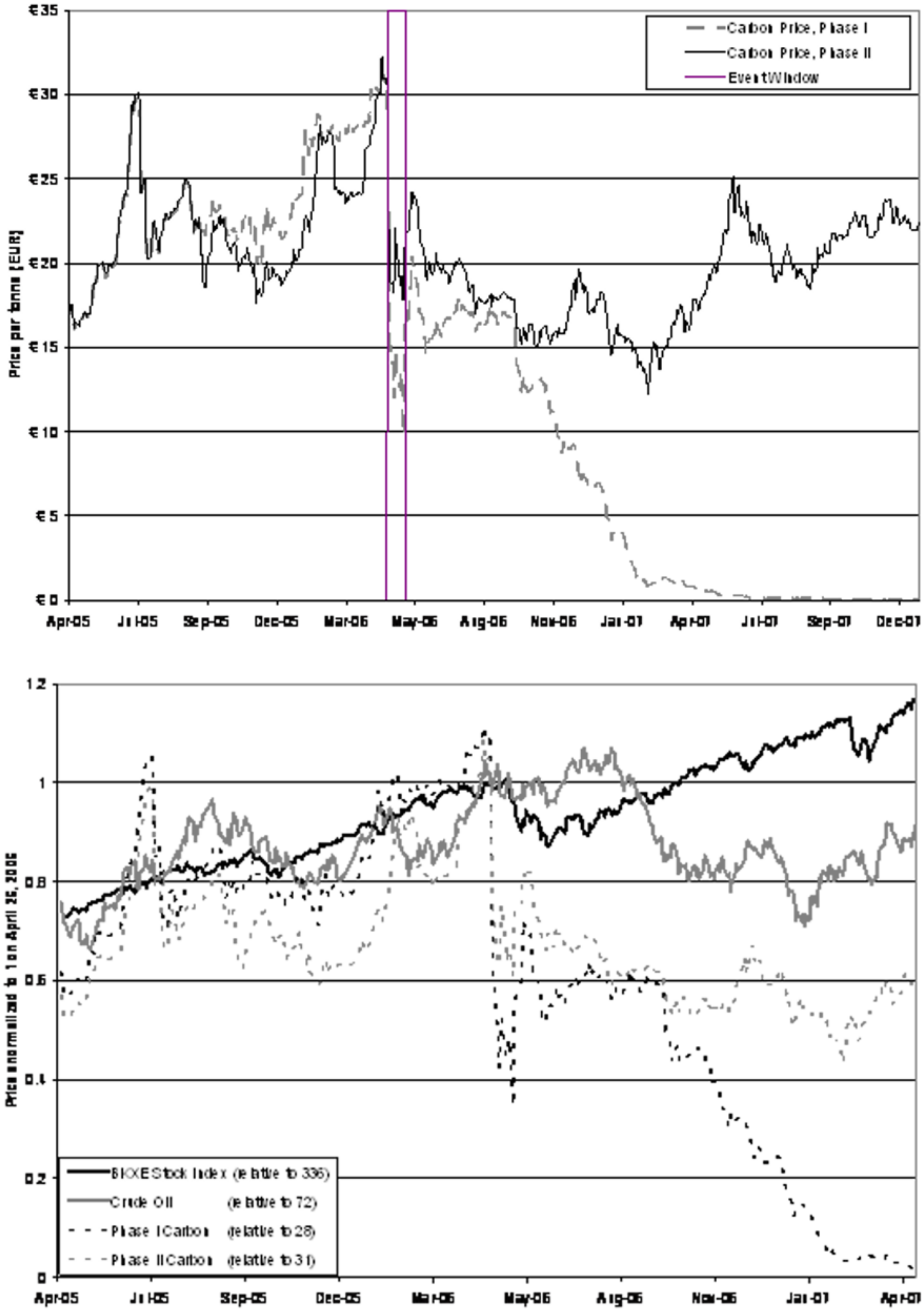


Figure 2.3: Most firms allowances are similar to emissions (Current subsample of 90 firms with emissions linked to stock market data)

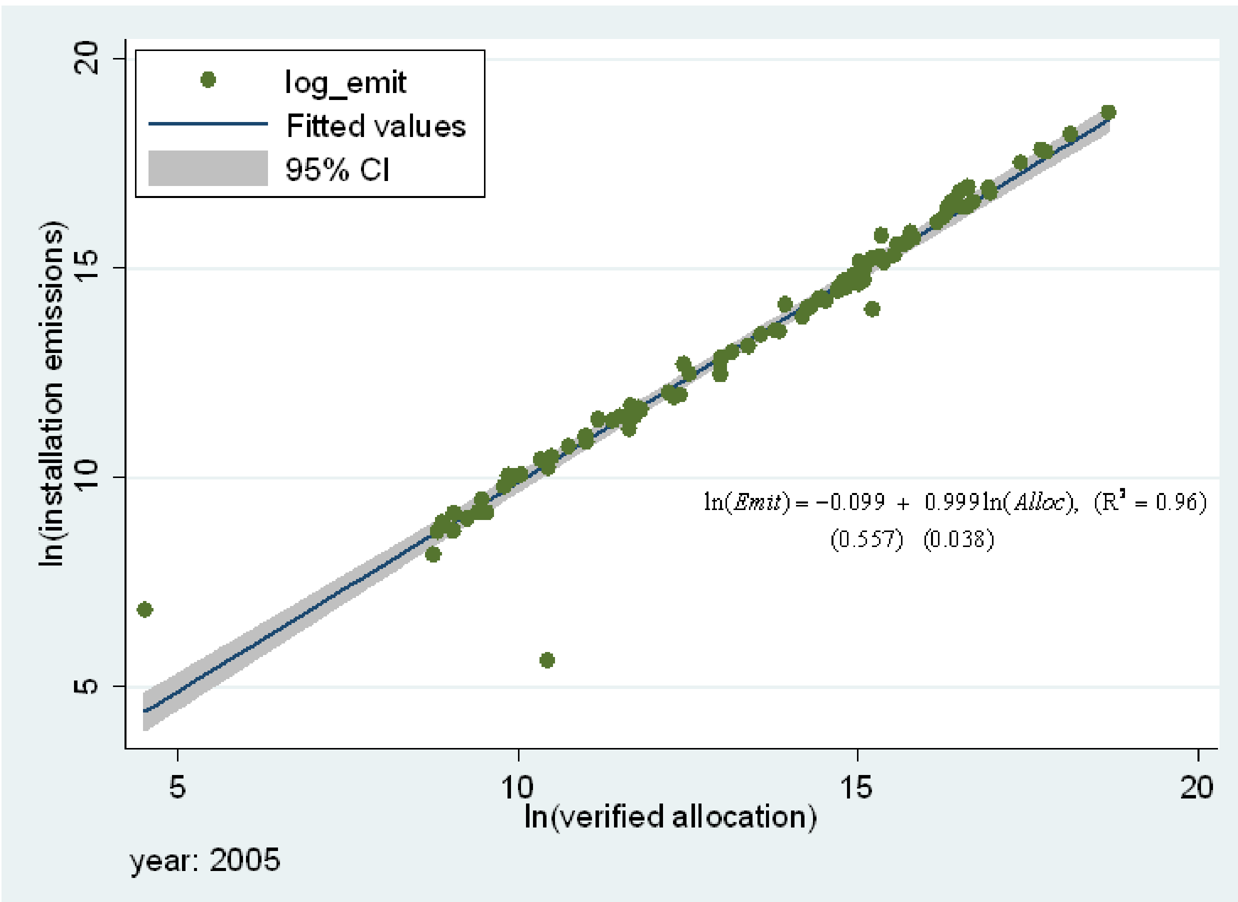


Table 2.1: Stock Market Cumulative Returns by Industry

		Cumulative Abnormal Returns			Cumulative Returns		
NAICS	Sector	Coef.	Std. Err.		Coef.	Std. Err.	
21	Mining & Oil/Gas Extraction	-0.0273	0.006	***	-0.0358	0.0057	***
22	Utilities	-0.0179	0.0056	***	-0.0211	0.0055	***
53	Real Estate & Rental	-0.0132	0.0048	***	-0.0169	0.005	***
23	Construction	-0.0115	0.0073		-0.0188	0.0073	**
72	Accommodation & Food Services	-0.0081	0.0058		-0.0117	0.0054	**
33	Manufacturing (Metals, Machinery)	-0.0068	0.0033	**	-0.0143	0.0032	***
45	Retail (General, Misc)	-0.0059	0.0022	***	-0.0098	0.0018	***
31	Food & Textiles	-0.0053	0.0034		-0.0096	0.0034	***
54	Professional, Scientific, & Technical Services	-0.0052	0.0059		-0.0119	0.0057	**
32	Manufacturing (Paper, Plastics)	-0.0032	0.0037		-0.0092	0.0037	**
99	Other	-0.0016	0.0088		-0.0083	0.0094	
48	Transportation	-0.0001	0.0051		-0.0052	0.0052	
51	Information	0.0002	0.0032		-0.0067	0.0032	**
52	Finance & Insurance	0.0008	0.0018		-0.0061	0.0018	***
44	Retail (Electronics, Gas, Health)	0.0019	0.0047		-0.0043	0.0043	
71	Arts, Entertainment, & Recreation	0.0043	0.0101		-0.0011	0.0099	
62	Health Care & Social Assistance	0.011	0.003	***	0.0059	0.0031	*
56	Administrative & Support	0.013	0.0064	**	0.0048	0.0066	
49	Couriers & Storage	0.0139	0.0068	**	0.009	0.0071	
42	Wholesale Trade	0.0166	0.0067	**	0.0098	0.0073	
	All Sectors	-0.0045	0.0015	***	-0.0108	0.0015	***

Notes: Significance is noted at the 10% (*), 5% (**) and 1% (***) levels. Standard errors are robust. There are 572 observations. Industry is by NAICS.

Table 2.2: Stock Market Cumulative Abnormal Returns for Firms in the Electricity Sector

Panel A: Firm Level Cumulative Abnormal Returns				
Stock Name	Event	Carbon per MWh	Carbon per Equity	MWh per Equity
Fortum	-0.088	0.214	0.265	1.236
Verbundgesellschaft	-0.086	0.252	0.941	3.729
British Energy Group	-0.071	0.108	1.117	10.365
EDF	-0.05	0.104	0.466	4.496
RWE (XET)	-0.045	0.909	3.049	3.355
A2A	-0.024	0.287	0.36	1.255
Atel Holding 'R'	-0.022	0.213		
DRAX Group	-0.019	1.046	3.854	3.684
United Utilities Group	-0.018			
EDP Energias de Portugal	-0.015	0.712	1.809	2.541
International Power	-0.012	0.611	2.084	3.414
Red Electrica de Espana	-0.005			
Scot.& Southern Energy	-0.004	0.819	1.92	2.344
ENEL	-0.003	0.501	1.466	2.926
National Grid	-0.001			
Terna	-0.001			
Union Fenosa	0.004	0.972	1.265	1.301
Schneider Electric	0.011			
Iberdrola	0.015	0.349	0.451	1.291
Public Power	0.052	0.982	8	8.146

Panel B. Correlations			
	Event	Carbon per MWh	Carbon per Equity
Carbon per MWh	0.593	1	
Carbon per Equity	0.58	0.689	1
MWh per Equity	-0.035	-0.091	0.476

Notes: NAICS 2211

Table 2.3: Tests of Net Permits at Firm Level

Panel A: All Industries (with NAICS3 Fixed Effects)								
	1		2		3		4	
Net Permits	-6.9							
	7.18							
Allocation			4.17				0.26	
			3.57				14.15	
Emissions					4.1		3.85	
					3.22		10.9	
Constant	F.E.		F.E.		F.E.		F.E.	
Panel B: Industries Net Short in Permits (Power Industry)								
	1		2		3		4	
Net Permits	-1.51							
	11.44							
Allocation			6.65	***			16.35	
			1.32				11.28	
Emissions					6.25	***	-9.45	
					1.5		11.14	
Constant	-0.022	**	-0.031	***	-0.031	***	-0.03	***
	0.008		0.007		0.008		0.008	
Panel C: Industries Net Long in Permits (with NAICS3 Fixed Effects)								
	1		2		3		4	
Net Permits	-17.11							
	29.18							
Allocation			-31.53	***			-6.73	
			6.81				20.24	
Emissions					-34.56	***	-27.64	
					6.79		17.58	
Constant	F.E.		F.E.		F.E.		F.E.	

Notes: Significance is noted at the 10% (*), 5% (**) and 1% (***) levels. Standard errors are robust. There are 90 observations in Panel A, 21 in Panel B, and 69 in Panel C. Firms in the power industry had an average net short position of 2.15 million while firms in other industries were on average net long by 282 thousand.

Table 2.4: Summary Statistics

Sector	N	Event Return	Dirty Output	Dirty Input	Trade Exposure	Market Cap
Mining & Oil/Gas Extraction	23	-2.73%	0.08	0.39	0.58	15,400
Utilities	28	-1.79%	0.97	0.04	n/a	18,900
Real Estate & Rental	21	-1.32%	n/a	0.14	n/a	4,870
Construction	28	-1.15%	0	0.1	n/a	6,680
Accommodation & Food Services	8	-0.81%	n/a	0.24	n/a	7,050
Manufacturing (Metals, Machinery)	100	-0.68%	0.05	0.22	0.52	9,950
Retail (General, Misc)	5	-0.59%	n/a	n/a	n/a	27,500
Food & Textiles	34	-0.53%	n/a	0.24	0.26	18,800
Professional, Scientific, & Technical Services	21	-0.52%	0	0.03	n/a	4,270
Manufacturing (Paper, Plastics)	65	-0.32%	0.1	0.3	0.35	30,100
Other	7	-0.16%	0.35	n/a	n/a	31,900
Transportation	23	-0.01%	0.03	0.09	n/a	6,610
Information	47	0.02%	n/a	0.05	n/a	18,600
Finance & Insurance	121	0.08%	n/a	0.01	n/a	22,100
Retail (Electronics, Gas, Health)	16	0.19%	0.09	n/a	n/a	11,900
Arts, Entertainment, & Recreation	4	0.43%	n/a	0.22	n/a	6,140
Health Care & Social Assistance	2	1.1%	n/a	0.09	n/a	6,830
Administrative & Support	8	1.3%	n/a	0.03	n/a	5,530
Couriers & Storage	3	1.39%	n/a	0.13	n/a	16,600
Wholesale Trade	8	1.66%	n/a	0.05	n/a	5,050

Notes: The table reports the sample mean for each two digit NAICS sector. Dirty Output is ratio of industrys emissions share to industrys equity share, Dirty Input is electricity costs over sales, Trade exposure is the ratio of the sum of imports and exports over the sum of imports and sales, and market cap is equity value in billions of US Dollars.

Table 2.5: Tests of Revenue Effects at Industry Level

	1	2	3	4	5	6	7
Dirty Output Indicator		-0.0034				-0.0016	-0.0005
		0.0029				0.0026	0.0023
DO Index		-0.0119 ***				-0.0131 ***	-0.0142 ***
		0.0031				0.0035	0.0039
Dirty Input Indicator			-0.001		-0.0002		-0.0004
			0.0043		0.0045		0.0045
DI Index			-0.0237 **		-0.052 **		-0.0526 *
			0.0109		0.0248		0.0268
Trade Exposure Indicator				-0.0044	0.0069	-0.0042	0.0068
				0.0038	0.0056	0.0038	0.0058
Trade Index				0.0023	-0.0108	0.0036	-0.0124
				0.0041	0.0118	0.0067	0.012
DI*Trade Index					0.0494		0.0564
					0.043		0.0506
DO*Trade Index						-0.0259	-0.0069
						0.0279	0.0326
Constant	-0.0045 ***	-0.0022	-0.0004	-0.0032	-0.0004	-0.0017	0.0012
	0.0015	0.0017	0.0039	0.002	0.0039	0.0019	0.0043

Notes: Significance is noted at the 10% (*), 5% (**) and 1% (***) levels. Standard errors are robust. There are 572 observations. Industry is by NAICS.

Table 2.6: Robustness to Big Event Window

	1	2	3	4	5	6	7
Dirty Output Indicator		-0.0034				0.0036	0.0027
		0.0074				0.0075	0.0072
DO Index		-0.012 *				-0.021 **	-0.0203 **
		0.0067				0.008	0.0079
Dirty Input Indicator			-0.0227		-0.0156		-0.0153
			0.0174		0.0175		0.0167
DI Index			-0.0139		-0.0977 **		-0.1012 **
			0.0169		0.0469		0.0485
Trade Exposure Indicator				-0.0033	0.0177	-0.005	0.0167
				0.009	0.0118	0.0088	0.0119
Trade Index				-0.0173	-0.0775 **	-0.0197	-0.08 **
				0.0247	0.0369	0.0223	0.036
DI*Trade Index					0.2392 **		0.2547 **
					0.1057		0.1201
DO*Trade Index						0.0248	-0.0397
						0.1539	0.1357
Constant	-0.0136 ***	-0.0113 **	0.0098	-0.0095 **	0.0098	-0.0086 *	0.0108
	0.0035	0.0046	0.0169	0.0041	0.0169	0.0044	0.0161

Notes: Significance is noted at the 10% (*), 5% (**) and 1% (***) levels. Standard errors are robust. There are 572 observations. Industry is by NAICS.

Table 2.7: Robustness to Including Debt-Equity Ratio Control

	1	2	3	4	5	6	7
Dirty Output Indicator		-0.0026				-0.0011	-0.0001
		0.0029				0.0025	0.0023
DO Index		-0.012 ***				-0.0129 ***	-0.014 ***
		0.0031				0.0036	0.0039
Dirty Input Indicator			-0.0016		-0.001		-0.001
			0.0044		0.0047		0.0046
DI Index			-0.0221 **		-0.0475 *		-0.0489 *
			0.011		0.0251		0.027
Trade Exposure Indicator				-0.0042	0.0062	-0.0042	0.0061
				0.0039	0.0057	0.0039	0.0058
Trade Index				0.0031	-0.0081	0.0044	-0.0101
				0.0041	0.0121	0.0067	0.0121
DI*Trade Index					0.0416		0.0496
					0.0436		0.0509
DO*Trade Index						-0.0276	-0.0076
						0.0277	0.0321
Debt-Equity Ratio	0.0011 ***	0.0009 **	0.0007 **	0.001 **	0.0007 *	0.0008 **	0.0006
	0.0003	0.0004	0.0003	0.0004	0.0004	0.0004	0.0004
Constant	-0.0052 ***	-0.0031 *	-0.0005	-0.0041 *	-0.0005	-0.0026	0.0009
	-0.0016	-0.0018	-0.0039	-0.0021	-0.0039	-0.0021	-0.0043

Notes: Significance is noted at the 10% (*), 5% (**) and 1% (***) levels. Standard errors are robust. There are 572 observations. Industry is by NAICS.

Table 2.8: Robustness to no CAPM (Cumulative returns)

	1	2	3	4	5	6	7
Dirty Output Indicator		-0.0044				-0.0025	-0.0013
		0.0031				0.0028	0.0025
DO Index		-0.0084 **				-0.0096 ***	-0.0108 ***
		0.0034				0.0035	0.0038
Dirty Input Indicator			-0.0011		-0.0005		-0.0008
			0.0037		0.0039		0.0038
DI Index			-0.0258 **		-0.0501 **		-0.0502 *
			0.0111		0.024		0.0258
Trade Exposure Indicator				-0.0032	0.0078	-0.0026	0.008
				0.004	0.0056	0.0038	0.0056
Trade Index				-0.0018	-0.0127		-0.0139
				0.0047	0.0118	0.0072	0.0122
DI*Trade Index					0.0406		0.0468
					0.0419		0.0486
DO*Trade Index						-0.031	-0.008
						0.0287	0.0326
Constant	-0.0108 ***	-0.0084 ***	-0.0063 *	-0.0093 ***	-0.0063 *	-0.0079 ***	-0.0048
	0.0015	0.0016	0.0033	0.0018	0.0033	0.0017	0.0036

Notes: Significance is noted at the 10% (*), 5% (**) and 1% (***) levels. Standard errors are robust. There are 572 observations. Industry is by NAICS.

Table 2.9: Spillovers to the United States (data from S&P 500)

	1	2	3	4	5	6	7
Dirty Output Indicator		-0.009				-0.0049	-0.0052
		0.0073				0.0066	0.0066
DO Index		0.0161 **				0.0139 *	0.0138 *
		0.0066				0.0075	0.0081
Dirty Input Indicator			-0.0037		-0.0011		-0.0023
			0.0055		0.0068		0.0075
DI Index			-0.0313 *		-0.0712		-0.0696
			0.0163		0.0486		0.0493
Trade Exposure Indicator				-0.002	0.0116	0.0022	0.0154
				0.0086	0.01	0.0081	0.0101
Trade Index				-0.0098	-0.0318	-0.0073	-0.0341 *
				0.013	0.0206	0.0101	0.0193
DI*Trade Index					0.0951		0.1216
					0.0858		0.0848
DO*Trade Index						-0.1632 *	-0.1593
						0.0927	0.1014
Constant	-0.006 **	-0.0041	0.0014	-0.0028	0.0014	-0.0035	0.0017
	0.0029	0.0034	0.0034	0.004	0.0035	0.0047	0.0035

Notes: Significance is noted at the 10% (*), 5% (**) and 1% (***) levels. Standard errors are robust. There are 572 observations. Industry is by NAICS.

Table 2.10: Counterfactual Event Study for April 2004

	1	2	3	4	5	6	7
Dirty Output Indicator		0.0009				-0.0027	-0.0023
		0.0044				0.0035	0.0033
DO Index		-0.004				0.0005	-0.0001
		0.0067				0.0045	0.0041
Dirty Input Indicator			-0.0055 **		-0.0069 ***		-0.0072 ***
			0.0026		0.0025		0.0027
DI Index			0.0052		-0.0159		-0.0148
			0.0146		0.0302		0.0307
Trade Exposure Indicator				0.0226 ***	0.0269 ***	0.0236 ***	0.0275 ***
				0.0065	0.0092	0.0067	0.0094
Trade Index				-0.0367 ***	-0.0354 *	0 ***	-0.0354 *
				0.0109	0.0182	0.0121	0.019
DI*Trade Index					-0.0066		-0.0047
					0.0476		0.0505
DO*Trade Index						-0.0248	-0.0041
						0.0301	0.0376
Constant	-0.0027	-0.0028	0.0019	-0.0052 ***	0.0019	-0.0046 ***	0.0027
	0.0018	0.0018	0.0014	0.0014	0.0014	0.0016	0.0019

Notes: Significance is noted at the 10% (*), 5% (**) and 1% (***) levels. Standard errors are robust. There are 531 observations. Industry is by NAICS.

Chapter 3

Classification, Detection and Consequences of Data Error: Evidence from the Human Development Index

Co-authored with Max Auffhammer and Hendrik Wolff. [1]

> Perhaps the greatest step forward that can be taken, even at short notice, is to insist that economic statistics be only published together with an estimate of their error. – Oskar Morgenstern, 1970

3.1 Introduction and Related Literature

This paper studies the Human Development Index (HDI), which has become one of the most widely used measures to communicate a country's development status. Compared to the Gross Domestic Product (GDP), the HDI is a broader measure of development, since it captures not only the level of income, but also incorporates measures of health and education (Srinivasan, 1994; Streeten, 1994; Anand and Sen, 2000). The United Nations Development Programme, which releases the HDI statistics, classifies each country into one of three categories: 'low human development' for HDI scores between 0.0 and 0.5, 'medium human development' for scores between 0.5 and 0.8 and 'high human development' for scores between 0.8 and 1.0.

Although these development categories were not originally designed to determine international relations, development aid or should imply any other legal consequences, today

[1] Published in 2011 in *The Economic Journal*. I thank the publisher, John Wiley and Sons, for permission to include this work in my Dissertation.

these three mutually exclusive categories are used in politics, academia, and the corporate world. In business relations, the categories have been used for international pricing purposes (Bate and Boateng, 2007). Since 2001 the pharmaceutical company Merck sells drugs at different prices with up to 90% discounts for countries that are classified as 'low', and 75% reductions for 'medium' countries (Petersen and Rother, 2001). Second, the HDI has been widely used in debates among development researchers and policy makers (Sen, 2000) and is actively invoked to structure discussions in development-political debates of both governmental and non-governmental organizations (NGOs) (Jahan, 2000; HDR, 1990 to 2006). For allocation of development aid, it is known that the government of Ireland puts a particular focus on countries categorized as 'low human development' (O'Neill, 2005). International climate accord designs following the expiring Kyoto Agreement have included a proposal for linking countries' abatement responsibilities according to their HDI (Hu, 2009). Thirdly, in economics, an extensive literature has studied the relationships between HDI rankings, economic growth, institutions, and other economic and social measures (Anand and Ravallion, 1993; Easterlin, 2000; Dasgupta, 2001). The conceptual underpinnings of the HDI can be found in the work by Amartya Sen (i.e. 1977, 1984, 1985, 1987). For a recent mathematical ethical rationalization of the HDI see Moreno-Ternero and Romer (2006). Oswald (1997), Blanchflower and Oswald (2005) and Leigh and Wolfers (2006) explore links between a happiness index and the HDI.[2]

Despite extensive use of the HDI statistics, the drastic changes in the distribution of HDI scores for developing countries, as displayed in Figure 3.1, have gone unnoticed in the academic and policy literature. When the HDI was first published in 1990, the cross country-distribution appeared to be approximately uniformly distributed between zero (least developed) and one (most developed). Today, however, the distribution is twin-peaked with two sharp spikes around the values of 0.5 and 0.8, which are the cut-off values for categorizing countries of 'low', 'medium' and 'high' human development.

In this paper, we investigate the role of data error on the published HDI and the consequences for its use in statistical analysis. We address these questions by exploiting (1) the originally published HDI time series, (2) the subindicator variables used to construct the HDI, (3) changes to the HDI formula, and (4) documented data revisions. We identify three sources of data error: measurement error due to data revisions, data error due to formula updating and misclassification due to inconsistent cut-off values. After isolating data revision error from error due to formula updates, we estimate country specific variances of the HDI scores. For example, the variance due to data revision for Bolivia represents the distribution of possible HDI values for Bolivia in a given year, which is solely created by updates to the data series. We show that the HDI contains data error standard deviations ranging from 0.03 (United States) to 0.11 (Niger), which is significant given the 0 to 1 scale.

[2]Other studies that specifically used the triple-bin classification include Kelley (1991), McGillivray (1991), Noorbakhsh (1998), Baliamoune (2004) in development economics, Mazumdar (2002), Noorbakhsh (2006) in macroeconomics, Hargittai (1998), Keiser et al. (2004) in communications and Guindon and Boisclair (2003) to analyze health outcomes across countries.

We find that the magnitude of the error variances is greater the lower the HDI rank, which is consistent with the quality of the statistical agencies improving with higher development. Likewise, country specific variances due to formula revisions are calculated. Mapping these cardinal noise measures onto the ordinal dimension, we find that 11%, 21% and 34% of all countries can be interpreted as currently misclassified in the development bins due to the three sources of data error, respectively.

We also investigate the ordinal rank error. Each year when the new HDI statistics are published, much public attention focuses on the relative rank of a country to its rank in prior years and to the rank position of competing countries. For example, when Canada lost the top HDI number 1 position in 2001, The National Post (3^{rd} of July, 2001) wrote: '*We're not No. 1! Canada drops in UN rankings... Prime Minister Jean Chretien often refers to the report in public statements and speeches...*'. Or, in 1998, when Pakistan (rank 138) bypassed India (rank 139), *The Tribune* (September 14^{th}, 1998) noted: '*Pak beat India, both lose!*'.[3] To investigate the reliability of such statements, we calculate each country's likelihood of deviation from its original published HDI rank. We find that on average the expected absolute deviation is nine rank positions. Furthermore, the average 95% confidence interval of our simulated HDI rank deviations ranges from -21 ranks to +20 ranks for the 2.5% percentile and the 97.5% percentile respectively. These calculations show that statements based on ordinal comparisons are to be interpreted with great care.

Our results have direct implications for the academic literature. First, there is a vast economic literature that uses the same country level data that are included in the construction of the HDI, namely purchasing power parity adjusted income (i.e. Rogoff, 1996), life expectancy (i.e. Acemoglu and Johnson, 2007) and the educational measures of literacy rate and school enrollment statistics (i.e. Krueger and Lindahl, 2001). We investigate the inherent noise characteristics for each of these variables separately by estimating country specific variances for the underlying variables—GDP per capita, school enrollment, literacy rate and life expectancy. We find that the variables of health and education exhibit particular large error variances in developing countries; in comparison income has a smaller error variance but among the three sub-indicators it reveals the largest updating bias. Second, the HDI has been used to analyze the evolution of the world's distribution of well being, to explore issues of inequality, polarization, foreign direct investment, development aid and to econometrically test various convergence hypotheses in macroeconomics. By replicating some of these studies and carrying out sensitivity analyses, we find that key parameters, such as estimated Gini coefficients and speed of convergence parameters, vary by up to 100% in their values solely due to the measurement error.

Our paper is related to the literature that discusses the challenges in accurately estimating national accounts and other aggregate statistics. Deaton and Heston (2008) provide an in depth analysis of the various factors that affect PPP. In their case, in order to eliminate

[3]Pakistan ranked 119 and 138 and India ranked 118 and 139 in 1997 and 1998 respectively. For an extended discussion about these and similar rank statements see Morse (2003).

differences in national price levels, GDP is combined with data by the International Comparison Program (ICP) but the ICP's methodologies are subject to various changes, (i.e. modifications of baskets, Laspeyres versus Paasche index, product quality adjustments). In discussing previous revisions of the methodologies, Deaton and Heston (2008) conclude that PPP data are '*not always suitable for the purposes to which they are put*'. Krueger and Lindahl (2001) study the relationship between economic growth and country level educational variables and discuss the direction of bias one might expect by using different variables. Other papers that characterize the noise in aggregate statistical data include Barro and Lee (2001) and de la Fuente and Domnech (2006) for educational measures, Dowrick and Quiggin (1997) and Neary (2004) for income based measures and Anderson (1999) for life expectancy. We add to this literature by systematically isolating the different sources of error into data based errors, formula based errors and cut-off value based errors. To our knowledge, this is the first paper to calculate country specific variance measures of the HDI, income, life expectancy, literacy rate, school enrollment, as well as to calculate indicators and probabilities of a country's misclassification.

The remainder of the paper is structured as follows. Section 3.2 describes the data. Section 3.3 outlines the framework and methods of measuring variances and misclassifications due to data revisions, formula changes and the threshold problem. Section 3.4 presents our results. Section 3.5 provides examples of how the HDI is used in various contexts and how errors can affects prior academic analysis. We conclude with policy recommendations in Section 3.6.

3.2 Data

The HDI is a composite indicator measuring a country's level of development along three dimensions: health, education and income. These dimensions are expressed as unit free and double bounded subindicators y_1, y_2, y_3, each taking values between zero and one. The subindicators themselves are functions of data $\boldsymbol{x}$ on primary and secondary school enrollment statistics, literacy rate, life expectancy and GDP per capita adjusted by purchasing power parity (PPP). Finally, the HDI is calculated as a simple average of the three ($k = 1, 2, 3$) subindicators, $\text{HDI} = 1/3\Sigma_k y_k(\boldsymbol{x})$, which is then used for ordinal and cardinal comparisons. The HDI is published annually in the Human Development Reports (HDR) by the United Nations Development Program (UNDP), which are available for the years 1990 to 2006 (HDR, 1990 to 2006).[4]

[4]The UNDP mainly draws the GDP data from the World Bank, the educational statistics are provided by UNESCO and life expectancy comes from the Population Division of the UN Department of Economic and Social Affairs. Since countries do not consistently provide data using the same methodologies, these data sets are complemented by data from the Penn World Tables as well as by UNDP's own estimates to impute missing values. See the technical appendices of the HDR (1990-2006) for details.

3.2.1 Original versus Revised Data

In our analysis we exploit the fact that the original historical data matrix $\boldsymbol{x}_t$ used by the UNDP in year t differs from the revised matrix $\boldsymbol{x}^R{}_t{}^s$ which includes updates between t and $s>t$. The original $\boldsymbol{x}_t$ is available for the years t =1999 to 2006, whereas the revised data $\boldsymbol{x}^R{}_t{}^s$ are available (i) for all years of the analyses, t = 1990 to 2006 and s = 2006 and (ii) for the HDI in t = 1975, the revised $\text{HDI}^R{}_{1975}{}^s$ is available for s = 1999, 2000,. . . , 2006. In this paper, $\boldsymbol{x}^R{}_t$ refers to the variables for year t kindly provided to the authors as of fall of 2006 by the UNDP office, except stated otherwise. $\boldsymbol{x}_t$ refers to the data that we hand-copied[5] from the t^{th} year Human Development Report (HDR, 1990 to 2006).

3.2.2 The HDI Formulas and Computation of Counterfactuals

Since 1990, the UNDP has made three major updates to the formula used to construct the HDI. For each year t and country i the HDI formula is given by

$$HDI_{it} = h_f(\boldsymbol{x}_{it}).$$

The formula $h$changed thrice as indexed by $f \in \{A, B, C\}$, which corresponds to the time periods 1990, 1995-1998 and 1999-2006, respectively. The three formulas are explained in the HDR technical appendices (1990 to 1999) of Jahan (2000) and in the appendix of this web based version.[6] We construct three 'counterfactuals' denoted by $h_A(\boldsymbol{x}^R{}_{it})$, $h_B(\boldsymbol{x}^R{}_{it})$, and $h_C(\boldsymbol{x}^R{}_{it})$. Hence, for the entire time series we recalculate what the HDI would have been if the alternate formulas had been in place, using the most recent available historical data on the subindicators. In the analysis we exploit exactly these differences between the 'original' HDI generated by the formula that was active at time t compared to the HDI generated by the other two formulas that were not active in that particular year t.

3.2.3 The Sample

We construct a balanced panel from 1990 to 2006. A country is included in our panel if it meets the following two conditions: (a) the country exists continuously between 1990 and 2006 (e.g., Croatia is dropped); (b) between the three revised subindicators and the countries' HDI as provided by the UNDP, the total sum of missing data points is less or equal to five. Furthermore, in some of our analysis we distinguish between industrialized and non-industrialized countries whereby the industrialized countries are defined as in Table 1.1 of HDR (1991). We impute any missing data points by linear interpolation. In this way we obtain a balanced panel 99 countries of which 76 are non-industrialized countries and 23 are industrialized countries.

[5]The data were hand-copied separately by two of the authors. Only after verifying that the two hand-copied data sets are 100% identical, we proceeded with the analysis.

[6]The web version is available at
http://faculty.washington.edu/hgwolff/EJOnlineWebVersionofHDI_Wolffetal2010.pdf.

3.3 Sources of Data Error and Methodology to Measure Data Uncertainty

This section provides a detailed discussion of the three sources of data error: measurement error due to data revisions (D), data noise due to formula updating (F) and misclassification due to inconsistent cut-off values (C), which we abbreviate by D, F and C. We propose a simple statistical framework to analyze these sources of error, which allows us to calculate country specific variances and confidence intervals and to simulate country specific probabilities of misclassification.

Before discussing the details of each source of error below, it is useful to illustrate when the different types of errors (D, F and C) enter into the construction of the HDI. The columns of Table 3.1 show the overall structure of the data and the rows display when each error category contributes to the data uncertainty, depending on the level of data analyzed. The first column shows that with respect to the primary data variables $\boldsymbol{x}$, the only source of error is due to data updating (D). For the subindicator functions $\boldsymbol{y}$, two sources of errors are identified. First, with respect to D, $\boldsymbol{y}$ is vulnerable because the data error of $\boldsymbol{x}$ is directly translated into $\boldsymbol{y}$ through the function $\boldsymbol{y}(\boldsymbol{x})$. Additionally, the nonlinear functions $\boldsymbol{y}(\boldsymbol{x})$ are subject to formula changes (F) over time. Similarly, the aggregate HDI measure is subject to D and F through HDI $= 1/3\Sigma_k y_k(\boldsymbol{x})$. The HDI development categories are subject to error type C. Finally, the three types of error can be calculated for any function of HDI, θ(HDI), *e.g.* Gini coefficients or regression parameters.

As we will make clear below, we calculate the three types of error independent of each other. Hence it is not the case that error measure F will implicitly include some data error D or vice versa. Only in Section 3.3.4 and 3.4.2 we show how the different type of errors add up and discuss the correlation structures among them.

What are the distinctions between these sources of errors? While the first error D is well known to econometricans as '*measurement error*', the changes to the data by F and C are due to subjective decisions by the data provider (here the UNDP). This subjective component changed over time and impacted the construction and relative importance of sub-variables of the HDI as well as the judgment on how to classify countries. Another distinction between D, F and C is that our first two types of errors, D and F, are *cardinal* in nature. This is in contrast to our third type of error, C, which is purely *ordinal* in nature in the sense that countries are either misclassified or not within the UN triple-bin classification system.

3.3.1 First Source of Data Error: Measurement error

To obtain the first measure of the randomness of the HDI data, we exploit the following exogenous changes to the data over time: The data $\boldsymbol{x}_t$ (as used by the UNDP for the HDR at year t) are in general not the same data as the UNDP publishes in year s for the same data year t. As revised statistics become available, the UNDP updates the original data matrix $\boldsymbol{x}_t$ at year s, $s \geq t$, which we then denote ${\boldsymbol{x}^R}_t{}^s$.

There are literally hundreds of reasons for data updates each year. The HDI draws their datasets from a multitude of domestic and international agencies (e.g. UNESCO, World Bank, Penn Tables). Often an agency may have data only for some subset of countries and some subset of years. The remaining data points are then filled by datasets from other agencies, and occasionally are interpolated by neighboring years or countries. The dozens of footnotes in the yearly HDR reports point to the institutions that changed data year by year. The complexity of the problem may be best illustrated with a specific example: since 1999, the UNDP publishes historic HDI scores going back until the year 1975, HDI_{1975}. Figure 3.2 displays HDI_{1975} scores as they are reported in each of the HDR reports from 1999 to 2006. In every year, between 1999 and 2006, substantial data revisions took place for the *same* 1975 HDI score. For example, while in 2000 Portugal was reported to have a historic HDI_{1975} of 0.73 (that was below the $\text{HDI}_{1975}$of Venezuela), by 2006 Portugal's HDI_{1975} increased to 0.79 and is now substantially above the 2006 reported HDI_{1975} of Venezuela. On average across all countries the HDI updating bias for the year 1975 can be calculated as 0.003 with a standard deviation of the updating error of $\sigma_{1975} = 0.012$. Given that the data updates took place after a quarter of a century, we consider 0.012 to be large. Instead, in a world of good data quality σ_{1975} should be close to zero. This implies that whenever an analyst uses UNDP data, the same analysis run at a later date will result in different estimates due to a changed data matrix. Hence, when the HDI is released in year t, the value must be understood as an inexact value subject to future data revisions. This problem is what we refer to as measurement error from data updating.

To parameterize this measurement error, assume that the relationship between the observed HDI score of country i and the true (but unknown) subindicators, denoted by ${y^*}_{itk}$, can be expressed as

$$HDI_{it} = 1/3\Sigma_k(y^*_{itk} + \varepsilon_{itk})$$

where ε_{itk} is orthogonal to ${y^*}_{itk}$ and is distributed with mean m_{itk} and country specific variance ${s^2}_{itk}$. The relationship between the observed HDI score of country i and the true HDI^* consequently is $\text{HDI}_{it} = {\text{HDI}_{it}}^* + e_{it}$ with e_{it} being the composite error term distributed with mean $1/3\Sigma_k m_{itk}$ and country specific variance ${\sigma^2}_i$ that is determined by the covariance structure of the measurement error of the subindicators in country i, $cov_i(\varepsilon_{tk})$.

Exploiting the original $\boldsymbol{x}_t$ and revised ${\boldsymbol{x}^R}_t$, we now are in the position to calculate country specific variances of the measurement error due to data updating (D) given by

$$\sigma^2_{D,i} = \Sigma_t(h_t(x_{it}) - h_t(x^R_{it}))^2/(T-1) \quad \forall t \in T \tag{3.1}$$

with h_t denoting the formula which was active at time t and $\boldsymbol{T} = A \cup B \cup C \backslash 2006$ is the union of the three time periods A, B, C, except for the last year of 2006. T denotes the number of elements in set $\boldsymbol{T}$. The variance of the data-updating measurement error is based on the difference between the original HDI as published in the HDR at year t, $h_t(\boldsymbol{x}_{it})$, and the reconstructed counterfactual HDI for year t using revised data ${\boldsymbol{x}_{it}}^R$ available to us today. To obtain a consistent estimate of the variance, we assume that $h_t({\boldsymbol{x}_{it}}^R)$ represents our currently best available estimate of $\mathrm{HDI}_{it}{}^*$ and discuss in our result the implications of this assumption.

Importantly note that we calculate $\sigma^2{}_{D,i}$ independently from error type F. Specifically, we disentangle D from F by constructing each pair of data $\{h_t(\boldsymbol{x}_{it}),\ h_t({\boldsymbol{x}_{it}}^R)\}_t\ \forall\ t \in \mathrm{T}$ to be conditional on the same HDI formula, namely the formula that was active at time t. (Instead, if one were using the pairs of data $\{h_t(\boldsymbol{x}_{it}),\ HDI^R\}$ as reported in the yearly UNDP reports, one would have erroneously incorporated error-type F into error type D).

3.3.2 Second Source: Changes in HDI Formula

Since its release in 1990, the HDI was often criticized with respect to its analytical framework and methodology (Desai, 1991; Kelley, 1991; McGillivray, 1991; Aturupane *et al.*, 1994; Noorbakhsh, 1998). The UNDP responded to this challenge by working with Nobel laureate Amartya Sen, Sudhir Anand, Paul Streeten and others to intellectually lead an effort to update the methodology and value judgments. As a result UNDP has made three major updates to the formula used to construct the HDI which are further discussed in Anand and Sen (1994, 1997, 2000), Jahan (2000), the technical appendices of the HDRs (1990 to 2006) and summarized in the appendix of the web based version paper. These three changes are clearly visible in the empirical distribution of the HDI displayed in Figure 3.3. In particular, different distributional characteristics occur for the sub-periods A (1990), B (1995-1998) and C (1999-2006) that correspond to the three formula regimes $h_A(\boldsymbol{x}_{it})$, $h_B(\boldsymbol{x}_{it})$, and $h_C(\boldsymbol{x}_{it})$, respectively.

We exploit this variation of the HDI scores across the counterfactual formulas to calculate country specific variances due to the formula (F) updates that is

$$\sigma^2{}_{F,i}= \Sigma_t\Sigma_g(h_g({\boldsymbol{x}_{it}}^R)-h_C({\boldsymbol{x}_{it}}^R))^2/(2T-1)\ \forall\ t \in \mathrm{T} \tag{3.2}$$

where g is the index to sum over the formula indices A, and B. The variance $\sigma^2{}_{F,i}$ is based on the country specific differences of the HDI generated by the most recent and improved formula h_C compared to the HDI counterfactuals generated by the other two formulas h_B and h_A. We do acknowledge that the formula revisions were undertaken to improve the HDI statistics and hence one interpretation of $\sigma^2{}_{F,i}$is to understand it as a measure of *historic* noise due to the formula updates. Alternatively, the country specific measures $\sigma^2{}_{F,i}$ can be interpreted as a *present* measure of noise, if the UNDP will similarly continue to change the formula in the future and the scores today would have to be understood as subject to those future formula revisions.

Note that we again isolate the error type F from the former error type D. Hence it is not the case that error-type F incorporates error-type D, and/or vice versa.[7]

[7]We achieve the independence because the function σ_F is defined conditional on the revised data x_R.

3.3.3 Third Source of Misclassification: Arbitrary Cutoff Values

In comparison to the *cardinal* measures of noise due to D and F, our third measure of error, C, is entirely *ordinal.* It is an error of misclassification due to the arbitrariness of the two cut-off values used to categorize countries into 'low', 'medium' and 'high' development countries. Despite the fact that changes made to the HDI formula did have considerable impacts on the empirical HDI distributions as displayed in Figure 3.3, the UNDP has decided to use the *same* cut-off values (0.5 and 0.8) since 1990. Since the original cutoff-values are supposed to distinguish three *qualities* of human development, with each formula change the UNDP could instead have adjusted the cut-off values in such a way that the new adjusted thresholds again reflect these same value judgments for the levels of quality. One possible procedure[8] to obtain revised threshold values—that are consistent with the initial 1990 value judgment of classifying quality *and* consistent with the entire history of formula changes—is as follows. In 1990, Morocco and Egypt were the two countries closest around the original cut off value of 0.5 (with HDI scores of 0.49 and 0.50, respectively). On the counterfactual distribution of formula h_C applied to 1990, these two countries take on the values 0.54 and 0.56. Taking the mean (0.55) provides the revised threshold for separating between the low and medium human development groups. Similarly we proceed with the cut off value 0.8 and obtain the revised value 0.70.

3.3.4 Overall Error Variance

So far, we have treated the two sources of errors D and F independently of each other. The user of the HDI statistics may, however, be also interested in having a sense of the "overall" error within the HDI database.[9] To this end, we calculate the country-specific overall cardinal error variance statistics as

$$\sigma^2(overall)_i = \sigma^2{}_{Di} + \sigma^2{}_{Fi} + 2cov(e_{Di}, e_{Fi})$$

which takes into account the covariance structure of the individual error contributions,

$$cov(e_{Di}, e_{Fi}) = \Sigma_t (e_{Dit} - m_{Di})(e_{Fit} - m_{Fi})/T$$

whereby $e_{Dit} = HDI_{it} - HDI^R{}_{it}$ and $e_{Fit} = \Sigma_f (h_f(\boldsymbol{x}_{it}{}^R) - h_C(\boldsymbol{x}_{it}{}^R))/2$. We can thus analyze how much each source of error (1) and (2) contributes to the overall level of error in the HDI database.

Hence all terms on the right hand side of σ_F are 'counterfactual' measures, what the HDI would have been if the revised data x_R had been already known in prior years under the different formula assumptions.

[8]Our procedure to choose the revised bin cutoffs is based upon the objective to maintain constant the initial (1990) value judgment by the UNDP, in the sense that the thresholds separate low from medium and medium from high developed countries. One referee suggested selecting those cutoff values which maximize the objective function to maintain the development category of as many countries as possible. This would lead to the revised thresholds values of 0.62 and 0.76.

[9]We thank the editor for providing the idea to aggregate errors.

3.3.5 Simulation 1: The expected number of misclassified countries

For the cardinal sources of data error, for each country we can calculate the probability of being misclassified. Given the parameterization of the measurement error as $\text{HDI}_{i2006}{}^*$ = HDI_{i2006}- e_{i2006} and assuming e_{i2006} ~ $N(0,\sigma^2{}_{.,i})$, normally distributed with mean zero and variance $\sigma^2{}_{.,i}$(as calculatedby $\sigma^2{}_{F,i}$, $\sigma^2{}_{D,i}$, and $\sigma^2(\text{overall})_{,i}$) we analytically calculate for each country the probability of being misclassified as

$$\begin{gathered}
\int_{0.5}^{1.0} p(\widehat{HDI}_i) d\widehat{HDI}_i \quad \forall is.t.HDI \in [0.0, 0.5) \\
\int_{0.0}^{0.5} p(\widehat{HDI}_i) d\widehat{HDI}_i + \int_{0.8}^{1.0} p(\widehat{HDI}_i) d\widehat{HDI}_i \quad \forall is.t.HDI \in [0.5, 0.8) \\
\int_{0.0}^{0.8} p(\widehat{HDI}_i) d\widehat{HDI}_i \quad \forall is.t.HDI \in [0.8, 1.0]
\end{gathered} \tag{3.3}$$

where $p()$ is the probability density function of the estimated HDI_i* distributions. Hence, for countries reported to be 'low development', we calculate the probability of being classified as a medium or a high development country; similarly, we proceed for the 'medium' and 'high' development countries. Finally, adding these integrals over all countries provides the expected number of misclassified countries.

3.3.6 Simulation 2: The expected number of deviation in HDI ranks

In addition to sorting countries into the three broad HDI categories of 'low', 'medium' and 'high', the UNDP statistics are used to produce league rankings of countries. We calculate the expected number of absolute deviations in rank by simulating (n = 1,. . . , 10,000) the 2006 HDI ranking. The simulated rankings are produced by calculating for every country i the simulated HDI as $\text{SimHDI}_{i,2006}$ = $\text{HDI}_{i,2006}$+ η_i with η_i distributed as mean zero and variance $\sigma^2(overall)_i$. Finally, after each nth simulation country *i's simulated* rank is recorded relative to its actual observed rank in 2006.

3.4 Results

3.4.1 Results with Respect to the Cardinal Errors of Data Updating and Formula Changes

If one followed Oskar Morgenstern's (1970) advice given in the introduction, an alternative way for UNDP to report HDI scores would be to report country specific noise measures. To do so, we display country specific standard errors in Table 3.2. With respect to the standard errors due to the measurement error of data updating (column 8), we find that $\sigma_{D,i}$

ranges between a minimum value of 0.004 (United States) and a maximum value of 0.069 (Syria), with an average value across all countries of 0.026. Given that the HDI is an average over three subindicators, whereby positive and negative deviations in the subindicators cancel out,[10] and given that the HDI is scaled from of 0 to 1, these standard deviations are large and significant.Figure 3.4 displays the relationship between the country specific measurement error due to the data revisions, $\sigma_{D,i}$ and the countries' HDI score (as of 2006). We note that more developed countries have smaller updating variances. Similarly column (3) displays the country specific data measurement errors due to formula updates $\sigma_{F,i}$, whose ranges on average are even higher compared to $\sigma_{D,i}$. We find the estimated $\sigma_{F,i}$ range between a minimum value of 0.034 and a maximum value of 0.127 with a world average standard deviation of 0.072.

Since the HDI is primarily used as an ordinal measure, we now turn to the impact of these cardinal measures on the ordinal dimension. Figure 3.5 displays the case of the "average" non-industrial country with HDI = 0.65 using the average standard deviation over *all* non-industrialized countries due to data revisions, σ_D=0.03 and due to formula updates σ_F=0.08. Figure 3.5 shows that substantial probability mass is spread over all three development categories. In Table 3.2, the category specific probabilities are displayed for all countries in columns 4-6, and 9-11 for the formula based error and data upgrading errors respectively. For example, as of 2006, South Africa, Mongolia, Syria, India, Honduras, Bolivia have non-zero probabilities of belonging to all three categories simultaneously. Even a high human development country, such as Costa Rica with HDI of 0.84, can still be a 'low' with 0.3% probability and yet be 'medium' to 37%. Finally, columns 7 and 12 display the total probability of a particular country being misclassified by using formula 3.3. The sum over these column probabilities show that currently, in expectation, 10.4 countries are misclassified due to data updating measurement error and 20.7 countries are misclassified due to formula updates; these numbers translate into, 11% and 21% of all countries being misclassified.

We interpret the misclassification of 11% due to data updating as *conservative* because ${\sigma^2}_{D,i}$ is just based on "short term" differences between $\boldsymbol{x}_t$ and ${\boldsymbol{x}^R}_t$, based on the years from 1990 to 2006. There also exists "long term" data updating error, which taking into account, that may increase ${\sigma^2}_D$ as $|\mathrm{HDI}_t \; {\mathrm{HDI}^R}_{t}{}^{s}|$ increases with s. While we cannot capture this long-term effect by formula 3.1 (due to the lack of published original data prior to the HDR of 1990), we illustrated the magnitude of such "long term" drift in Figure 3.2.

[10]The correlation between the three subindicator error terms $\varepsilon_i tk$, $k \in \{1,2,3\}$ is close to zero and can be viewed as distributed approximately independently. Hence the average standard deviation of the subindicator errors s_k^2 must be larger in magnitude, compared to the standard deviation of the HDI, $\sigma_{D,i}$. Section 3.4.4 confirms this by analyzing the compound error term.

3.4.2 Overall Cardinal Error and Rank Simulations

The typical user of the HDI statistics may not be concerned about the individual error statistics ${\sigma^2}_D$ and ${\sigma^2}_F$ if they are calculated independently of each other, but the researcher may be more interested in obtaining a sense of the overall error in the data. For this purpose we calculate country specific overall cardinal error statistics $\sigma^2(\text{overall})_i$ and find that the world average of these measures $\sigma^2(\text{overall}) = \Sigma_i \sigma^2(\text{overall})_i/N$ equals to 0.007, compared to ${\sigma^2}_D = 0.001$ and ${\sigma^2}_F = 0.006$. Furthermore we find that all country specific covariance terms *cov*(e_{Di},e_{Fi}) are relatively small (all correlation coefficients are smaller than 0.06 in absolute value) which implies that the updating error is not linearly correlated with the formula error. This implies that 86% of the total HDI variance is contributed by the formula error and 14% by the measurement error due to data updating.[11] By using the same methodology as in Section 3.3.5, we calculate the "overall" expected number of countries misclassified as 22.9. The country specific overall variance statistics are given in column 3 of Table 3.3.

Moreover, column 1 of Table 3.3 displays the country specific expected absolute value of rank displacements based upon the rank of the country's HDI in 2006. Worldwide, the average country is displaced by about nine ranks. This average absolute displacement obscures the direction of rank displacement and the uncertainty over rank displacements. To this end, Figure 3.6 displays the average rank displacement over 10,000 simulations as a function of the countries' 2006 HDI score along with the 95% confidence intervals. The confidence intervals are large, leading to an average deviation of -21 ranks and +20 ranks for the 2.5% percentile and the 97.5% percentile respectively. Figure 3.6 also shows that countries with a low initial 2006 rank (low HDI score) do on average better in the simulated rank statistics and countries with an initial high HDI in 2006 are more likely to lose ranks in the simulations.

3.4.3 Results with Respect to the Cutoff Value Problem

Our third measure of misclassification is due to the non-adjustment problem of the cut-off values 0.5 and 0.8 that the UN uses to classify countries as low, medium and high human developed countries. If the UNDP had adjusted the cut-off values in a manner consistent with the 1990 classification, since 1999 (the year of the last formula update), the thresholds should be at the values 0.55 and 0.70, as opposed to 0.5 and 0.8. This lack of adjustment of the cutoff values results in 34% of the countries being misclassified today.[12] Among all developing countries the percentage of misclassification is even higher: 45%. With such a high percentage statements such as '*over the last decade x% of African countries successfully moved from the 'low' to the 'medium' human development category*'—

[11] We calculate the % contribution of the jth cardinal source of error to the overall error as $\frac{\sigma_i^2}{\sigma_{overall,i}^2}$. This calculation is hence net of the covariance of the two error sources. The covariance terms can essentially be neglected due to the fact these are small in magnitude.

[12] The percentage of countries misclassified is calculated as the number of countries that have HDI scores in the ranges [0.5, 0.55) and [0.70, 0.8) divided by the total number of countries in our sample (99).

as expressed in numerous policy papers and news reports (United Nations 1996, People's Daily 2001, Daily Times 2005)—become useless at best, if not blatantly misleading. The listing of the misclassified countries due to this source of error is provided in Table 3.4.

3.4.4 Measurement Error with Respect to the Underlying Variables of the HDI

Thus far, we analyzed the data error of the HDI. Since the same variables used to construct the HDI serve as key data in many academic studies as well as inputs to many other international comparative statistics, it is worthwhile analyzing the subindicators $\boldsymbol{y}$ pertaining to health, education and income in more detail.

The first four columns of Table 3.5 display summary statistics of the overall HDI updating error, e, and the vector of subindicator updating errors, ε, for our sample of 76 non-industrialized countries. In general, the standard deviations of the health and education indexes are larger than the standard deviations of the income statistics. It is interesting to note, however, that the main driver for the HDI upward bias stems from the change to the income index (m_{income}=0.01).[13] Instead, the errors on the health and the education indices show distributions that are centered around zero. Note, that the min/max columns in Table 3.5 reveal some enormous changes; the income index changed by 15% (Sudan and Chad) and the education index even by 25% (Mongolia) on the total scale from 0 to 1.

One may ask whether the three subindicator updating errors are correlated. An analysis of the year-by-year correlation matrices of the errors does not show any systematic co-movement, as the correlation coefficients are close to zero in all years. This suggests that the statistical adjustments on the three dimensions are independent of each other and indicates that the respective national statistical offices responsible for health, education, and income statistics have no systematic contemporaneous responses. Furthermore, statistical independence of the three subindicator error variables ε_k implies that their errors must be on average larger than the variance of the HDI error e, which is confirmed by Table 3.5. Hence, while the three subindicator errors offset each other with respect to the HDI,[14] when working with the variables of education, income and health, one faces even larger data error.

To analyze the drivers of the HDI data error in more detail, we calculate country specific noise measures due to data revisions with respect to the underlying variables, $\boldsymbol{x}$. Table 3.6 reports country specific standard errors calculated as the country specific standard deviation $\sigma(x_n)_i$ (computed analogously to (2) by exploiting the 2006 data revision of x_{nit} for $t = 1999$ to 2005). In order to obtain a sense of the relative magnitude of the errors in each variable, we divide the standard deviations by the level corresponding variable in the year 2006, x_{ni2006}

[13] Statistically, this upward bias with a standard deviation of 0.02 is not significantly different from zero.

[14] Under the assumption of independence, the standard deviation for the composite HDI error, e, is given by std(e)=SQRT[$(\Sigma_k s_k^2/9)$], which, after replacing s_k, equals to std(e)= 0.014. The estimated standard deviation of the HDI measurement error by formula (3.1) applied to period C is 0.015, hence very close to std(e), confirming this theoretical result of independence.

and display the resulting relative standard errors in Figure 3.7. Adult literacy rate, GDP and the gross enrollment ratio contribute most to the updating error of the HDI. In contrast, life expectancy is revised much less. As is clearly recognizable in Figure 3.7, we find that the more highly developed the country the smaller its measurement error due to data updating.

3.5 Discussion of the results

Given that the HDI is subject to a considerable amount of measurement error, the use of the HDI and its triple bin classification system can lead to serious interpretability problems. We now investigate the consequences of these three sources of errors by replicating prior studies and uses of the HDI, with each of the analysis being uniquely linked to our three sources of errors.

3.5.1 The HDI as a definitional measure

While there does not exist a standardized definition of the term "developing country", the definition is often linked to the HDI, as being a country with low to moderate development status. In fact, often scientific studies have been explicitly using the HDI system to identify a set of developing countries (i.e. Noorbakhsh, 2006; Varenne, 2007; Lauber and Roessler, 2007; Alvan, 2009). Leading online dictionaries do refer to the HDI in order to define the term "developing country" (Wikipedia, 2008; Babylon, 2009; SearchWiki, 2009). Here it is common to differentiate development status by using three different colors. In Figure 3.8, we recreate such a map by displaying the HDI scores for 2006. To demonstrate the impact of misclassification in our sample, we reclassify the countries using the updated thresholds of 0.55 and 0.70 as discussed in Section 3.3.3. The visual impact of this reclassification is striking, especially in South America, Southeast Asia and Africa. This misclassification is particularly problematic, if organizations/institutions use these categories to design particular policies or rules.

3.5.2 The HDI and Foreign Development Aid:

Although, to our knowledge, the HDI is not formally used by any development agency as the sole index used to determine the distribution of development funds, there are clear indications that the HDI plays a significant role in governmental institutions' and NGOs' decisions for foreign aid allocation.[15] In 2000, the Deputy Director of the UNDP exemplified this debate by stating:

[15]For a related discussion see Alesina and Dollar (2000), Alesina and Weder (2002), Arcelus et al. (2005), Bandyopadhyay and Wall (2006), Easterly et al. (2004).

'At the global level, issues are now being explored as to whether bilateral aid can be allocated on the basis of HDI, or the core funds of multilateral agencies can be based on the index [...]' (p. 10, Jahan, 2000).
In fact, 'charity scorecards' are increasingly used as a tool for helping individuals decide which countries to donate money to. Here the HDI can be used to construct such a score. For example, on the homepage of http://www.charityscorecard.org/ a world map of HDI scores is displayed. The use of the HDI in this context may explicitly and implicitly steer users to "misclassified countries". Further, the triple bin classification is often used for report writing purposes to describe donor activities by governmental organizations (United Nations, 1996; HDR, 2001 to 2007) and non-governmental organizations. For example, Geneva Global (2007), which holds investments of 60 million client dollars in development projects, structures its funds according to the three HDI categories. For each year, the United Nations (HDR 2001 to 2006) analyzes the newest data on development aid as a function of the three human development categories. Drawing on these HDR statistics, Table 3.7 summarizes that across all years countries in the 'low' category obtained 3.4 times the official development assistance (ODA) per capita as compared to the medium development countries, which we do not claim is a causal effect but rather an interesting correlation.

3.5.3 Use of the HDI statistics in the academic literature

The HDI has been increasingly employed in the academic literature to describe the evolution of the world's "welfare" distribution in terms of various measures of inequality, such as the Gini coefficient, and to discuss the path of polarization, *e.g.* Pillarisetti (1997), Ogwang (2000), Mazumdar (2002), Noorbakhsh (2006), Prados de la Escosura (2007). The results published in these studies can differ greatly depending on which year the researcher collected the data in. To illustrate, Figure 3.9 displays HDI Gini coefficients using the formulas h_A, h_B and h_C for data covering 1975 to 2005 in five years intervals. The values produced by formula h_A are 25% to 50% *higher* and the time trend *steeper* compared to the time series generated by formula h_C. This substantial difference would lead to different conclusions or policy recommendations by the analyst. For a recent discussion on the relevance of levels and gradients of Gini estimates see for example Sala-i-Martin (2006) and Prados de la Escosura (2007).

We find that a number of recent studies are sensitive to random selection of countries that is due to the "arbitrariness" of the cut-off values: For example in the macroeconomic literature, Mazumdar (2002) and Noorbakhsh (2006) use the triple bins to analyze the existence of convergence clubs (Quah 1996) by testing the beta and the sigma conditional convergence hypothesis, originally discussed in Barro and Sala-i-Martin (1992). In particular, Noorbakhsh (2006) runs beta-convergence regressions of the form

$$ln(hdi_{it+T}/hdi_{it})/T = \alpha + \beta ln(hdi_{it}) + \epsilon_{it} \qquad (3.4)$$

conditional on the country belonging to the 'low' development bin. The dependent variable is the annualized growth of the HDI variable for country i over the period t to $t+T$ and hdi_{it} is the ratio of HDI in the i^{th} country to the average for the sample.[16] The regression is then repeated for the bins 'medium' and 'high' and the comparison of the β estimates is used to analyze the existence of convergence clubs.

To illustrate the consequences of the random selection, we first rerun the convergence regression 3.4 conditional on the HDI being in the interval $A_0 = [0.5, 0.8)$ as specified in Noorbakhsh (2006, p. 10, table 3). Then we perform the same regression with the adjusted cut-off values in the set $A_1 = [0.55, 0.70)$. The results are displayed in Table 3.8. Comparing the main parameter of interest, β, the estimate of the second regression is about 100% off the first regression implying a much faster speed of convergence.[17] This demonstrates that results based on the reported HDI can be very sensitive to changes of the HDI triple bin classification system.

3.5.4 Implications of the results in statistical analysis

Econometrically speaking, the average error measures $\sigma^2{}_D$ and $\sigma^2{}_F$ calculated in Section 3.4.1 imply that there is a 3% and 14% downward attenuation bias in a ordinary least squares (OLS) regression, $y = \beta_1 + \beta_2\text{HDI}^* + \epsilon$, if the observed HDI—instead of the "true" (but unknown) HDI*—is used as the regressor (for any variable y of interest). The bias of the OLS estimate b_2 is given by[18]

$$\text{plim } b^D{}_2 = [1\text{-}\sigma^2{}_D/(\sigma^2{}_D+\sigma^2{}_{HDI*})]\beta_2 \approx 0.97\beta_2$$

and

$$\text{plim } b^F{}_2 = [1\text{-}\sigma^2{}_F/(\sigma^2{}_F+\sigma^2{}_{HDI*})]\beta_2 \approx 0.86\beta_2.$$

This is important since in many econometric cross-country studies the HDI is used as a regressor, i.e. Globerman and Shapiro (2002), Mazumdar (2002), Sanyal and Samanta (2004), Neumayer (2003), Noorbakhsh (2006), Leigh and Wolfers (2006). This is even more crucial when working with the individual subindicator variables, since (as shown in Section 3.4.4) their average standard deviation of the measurement error is larger than the error of the HDI.

[16] A value of β in the range of (-1, 0) would imply β-convergence of the countries in the sample. A β of zero means no convergence and a positive value for β indicates divergence, with the speed of convergence/divergence the higher the absolute value of β.

[17] Note that the two β estimates are statistically significant with t values of -6.74 and -.4.59 for the sample of countries in A_0 and in A_1, respectively. We reject at the 1% significance level the hypothesis of uniform convergence in A_0 and in A_1 based on the Wald test examining whether β_1 is different from β_0, based on the pooled sample with appropriate interaction terms, with standard errors clustered by country.

[18] σ^2_{HDI*} is approximated by the empirical analogue of the 2006 HDI scores, $\hat{\sigma}^2_{HDI*}= 0.036$.

3.6 Conclusions

This paper identifies three sources of HDI data error and we make the following empirical contributions. First, we calculate country specific noise measures due to measurement error, formula choice and inconsistencies in the cut-off values. We find that the HDI statistics contain a substantial amount of noise on the order of 0.01 to 0.11 standard deviations. In analyzing the sources of the updating error we calculate country specific variances of GDP per capita, literacy rate, educational enrollment and life expectancy and we calculate the interdependence between these measures. We find that in general the higher the development status of a country, the more precise are the reported data. Second, we calculate the misclassification measures with respect to these three sources of data error by simulating the probabilities of being misclassified and sensitivity analysis of the cut-off values. We find that up to 45% of the developing countries are misclassified due to the failure to update the cutoff values. The discrete classification system is vulnerable when many countries are close to the thresholds, as is the case in the most recent years. Third, we discuss various empirical examples from the prior macroeconomic/development literature where the HDI has been employed and find that its use is problematic. Key parameters vary by up to 100% in their values. Although there may be certain benefits for the UN and charities for using a triple-bin classification system—bins are likely to improve publicity for the HDI and may hence help with more efficient internal organization of aid institutions—our results raise serious concerns about the system. We suggest that the United Nations should discontinue the practice of classifying countries into these triple bins because in our view the two cut-off values are arbitrary, can provide incentives for strategic behavior in reporting official statistics, and have the potential to misguide politicians, investors, charity donators and the public at large.

This paper did not investigate the drivers of why in the early years of the HDI—when its political role was still uncertain—the distribution as displayed in Figure 3.1 looked so different from today's. However, we caution governments, private investors, donor organizations and users of the charity scorecards not to take the triple bin system as a tool for international negotiations (Hu, 2009), foreign direct investments (Arcelus *et al.*, 2005), pricing (Bate and Boateng, 2007), or the allocation of foreign aid (Jahan, 2000; Neumayer, 2003). Such politically sensitive uses of the HDI might potentially provide perverse incentives for a country to manipulate the subindicator variables, if it has realized the comparative advantage of a 0.49 HDI score vs. a 0.51 score. In fact, announcements such as the statement by Jahan (2000) (discussed in Section 3.5.2) might have just created these incentives. We quote Oskar Morgenstern (1970):

'Governments, too are not free from falsifying statistics. This occurs, for example, when they are bargaining with other governments and wish to obtain strategic advantages or feel impelled to bluff [...]. A special study of these falsified, suppressed, and misrepresented government statistics is greatly needed and should be made.'

3.7 Tables, Figures, and Data Appendix

Table 3.1: Structure of the Type of Errors for Different Levels of Data Aggregation

Type of Error	$\boldsymbol{x}$	$\boldsymbol{y}(\boldsymbol{x})$	$HDI(\boldsymbol{y})$	$\theta(HDI)$
Data revisions D	$\surd$	$\surd$	$\surd$	$\surd$
Formula updates F		$\surd$	$\surd$	$\surd$
Cut-off value C			$\surd$	$\surd$

Note: For each column we indicate by the symbol $\surd$ which type of data error can affect the particular variable displayed in the column. $\boldsymbol{x}$ refers to the raw variables, $\boldsymbol{y}(\boldsymbol{x})$ to the subindicators which are functions of $\boldsymbol{x}$, HDI is a function of the $\boldsymbol{y}$ and $\theta(HDI)$ refers to any parameter of interest (i.e. Gini coefficient) that is calculated as a function of one or multiple HDI values.

Table 3.2: Country i specific standard deviations and probabilities of belonging to development category j

			Measures based on formula updates (F)					Measures based on measurement error due to data revisions (D)				
Country i and 2006 reported human development status		2006 HDI	$\sigma_{F,i}$	Pr{i='low'}	Prob{i='mid'}	Prob{i='high'}	Prob{i=mis-classified}	$\sigma_{D,i}$	Pr{i='low'}	Prob{i='mid'}	Prob{i='high'}	Prob{i=mis-classified}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Niger	low	0.31	0.13	93.1	6.9	0.0	6.9	0.03	100.0	0.0	0.0	0.0
Mali	low	0.34	0.12	91.9	8.1	0.0	8.1	0.03	100.0	0.0	0.0	0.0
Burkina Faso	low	0.34	0.11	92.3	7.7	0.0	7.7	0.02	100.0	0.0	0.0	0.0
Chad	low	0.37	0.11	89.3	10.7	0.0	10.7	0.04	100.0	0.0	0.0	0.0
Ethiopia	low	0.37	0.11	88.3	11.7	0.0	11.7	0.03	100.0	0.0	0.0	0.0
Burundi	low	0.38	0.11	85.4	14.6	0.0	14.6	0.02	100.0	0.0	0.0	0.0
Mozambique	low	0.39	0.11	83.4	16.6	0.0	16.6	0.03	100.0	0.0	0.0	0.0
Malawi	low	0.40	0.13	78.7	21.2	0.1	21.3	0.01	100.0	0.0	0.0	0.0
Zambia	low	0.41	0.09	86.2	13.8	0.0	13.8	0.04	98.5	1.5	0.0	1.5
Cote d'Ivoire	low	0.42	0.09	82.1	17.9	0.0	17.9	0.02	100.0	0.0	0.0	0.0
Benin	low	0.43	0.10	76.8	23.2	0.0	23.2	0.03	99.0	1.0	0.0	1.0
Tanzania	low	0.43	0.08	80.1	19.9	0.0	19.9	0.02	99.8	0.2	0.0	0.2
Nigeria	low	0.45	0.10	69.0	31.0	0.0	31.0	0.05	87.5	12.5	0.0	12.5
Rwanda	low	0.45	0.08	72.4	27.6	0.0	27.6	0.05	86.0	14.0	0.0	14.0
Senegal	low	0.46	0.08	68.5	31.5	0.0	31.5	0.02	99.5	0.5	0.0	0.5
Mauritania	low	0.49	0.08	56.6	43.4	0.0	43.4	0.03	66.5	33.5	0.0	33.5
Kenya	low	0.49	0.08	54.7	45.3	0.0	45.3	0.02	64.3	35.7	0.0	35.7
Zimbabwe	low	0.49	0.05	56.6	43.4	0.0	43.4	0.03	62.3	37.7	0.0	37.7
Lesotho	low	0.49	0.06	53.7	46.3	0.0	46.3	0.03	59.5	40.5	0.0	40.5
Togo	low	0.50	0.08	52.5	47.5	0.0	47.5	0.04	55.0	45.0	0.0	45.0
Uganda	med	0.50	0.10	49.2	50.7	0.1	49.3	0.02	46.2	53.8	0.0	46.2
Cameroon	med	0.51	0.07	46.6	53.4	0.0	46.6	0.04	44.6	55.4	0.0	44.6
Madagascar	med	0.51	0.08	45.4	54.6	0.0	45.4	0.03	39.3	60.7	0.0	39.3

continued on next page

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sudan	med	0.52	0.07	41.1	58.9	0.0	41.1	0.03	32.2	67.8	0.0	32.2
Congo	med	0.52	0.08	39.9	60.1	0.0	39.9	0.05	35.3	64.7	0.0	35.3
Papua New Guinea	med	0.52	0.06	33.9	66.1	0.0	33.9	0.04	27.7	72.3	0.0	27.7
Nepal	med	0.53	0.09	38.0	61.9	0.1	38.1	0.02	10.4	89.6	0.0	10.4
Bangladesh	med	0.53	0.08	35.3	64.7	0.0	35.3	0.02	7.3	92.7	0.0	7.3
Ghana	med	0.53	0.06	30.8	69.2	0.0	30.8	0.04	20.5	79.5	0.0	20.5
Pakistan	med	0.54	0.07	27.5	72.5	0.0	27.5	0.03	10.7	89.3	0.0	10.7
Lao People's Dem. R.	med	0.55	0.08	25.2	74.7	0.1	25.3	0.06	18.7	81.3	0.0	18.7
Botswana	med	0.57	0.05	8.3	91.7	0.0	8.3	0.04	3.4	96.6	0.0	3.4
India	med	0.61	0.06	3.0	96.9	0.1	3.1	0.01	0.0	100.0	0.0	0.0
Morocco	med	0.64	0.04	0.0	100.0	0.0	0.0	0.02	0.0	100.0	0.0	0.0
South Africa	med	0.65	0.05	0.2	99.5	0.3	0.5	0.07	1.3	97.2	1.6	2.8
Guatemala	med	0.67	0.04	0.0	99.9	0.1	0.1	0.02	0.0	100.0	0.0	0.0
Honduras	med	0.68	0.06	0.1	97.9	2.0	2.1	0.02	0.0	100.0	0.0	0.0
Mongolia	med	0.69	0.07	0.3	94.1	5.6	5.9	0.06	0.1	96.1	3.9	3.9
Bolivia	med	0.69	0.06	0.0	97.1	2.8	2.9	0.02	0.0	100.0	0.0	0.0
Nicaragua	med	0.70	0.05	0.0	98.6	1.4	1.4	0.04	0.0	99.2	0.8	0.8
Egypt	med	0.70	0.04	0.0	99.5	0.5	0.5	0.04	0.0	99.7	0.3	0.3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Viet Nam	med	0.71	0.07	0.2	88.9	10.9	11.1	0.02	0.0	100.0	0.0	0.0
Indonesia	med	0.71	0.06	0.0	93.9	6.0	6.1	0.03	0.0	99.9	0.1	0.1
Syrian Arab Republic	med	0.72	0.06	0.0	92.0	7.9	8.0	0.07	0.1	88.6	11.3	11.4
Jamaica	med	0.72	0.06	0.0	88.0	11.9	12.0	0.02	0.0	100.0	0.0	0.0
Algeria	med	0.73	0.04	0.0	95.6	4.4	4.4	0.04	0.0	97.5	2.5	2.5
El Salvador	med	0.73	0.05	0.0	90.3	9.7	9.7	0.05	0.0	91.1	8.9	8.9
Iran, Islamic Rep. of	med	0.75	0.05	0.0	84.6	15.4	15.4	0.03	0.0	98.1	1.9	1.9
Dominican Republic	med	0.75	0.06	0.0	80.0	20.0	20.0	0.02	0.0	99.9	0.1	0.1
Sri Lanka	med	0.76	0.08	0.0	72.3	27.7	27.7	0.02	0.0	96.6	3.4	3.4
Paraguay	med	0.76	0.07	0.0	72.1	27.9	27.9	0.03	0.0	93.0	7.0	7.0
Turkey	med	0.76	0.07	0.0	72.0	28.0	28.0	0.01	0.0	99.9	0.1	0.1
Jordan	med	0.76	0.07	0.0	73.1	26.9	26.9	0.03	0.0	89.1	10.9	10.9
Tunisia	med	0.76	0.06	0.0	76.2	23.8	23.8	0.02	0.0	96.2	3.8	3.8

continued on next page

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Philippines	med	0.76	0.06	0.0	73.8	26.2	26.2	0.03	0.0	90.6	9.4	9.4
Peru	med	0.77	0.05	0.0	74.1	25.9	25.9	0.02	0.0	96.9	3.1	3.1
China	med	0.77	0.06	0.0	69.7	30.3	30.3	0.02	0.0	94.7	5.3	5.3
Lebanon	med	0.77	0.06	0.0	68.1	31.9	31.9	0.04	0.0	74.9	25.1	25.1
Saudi Arabia	med	0.78	0.07	0.0	63.1	36.9	36.9	0.02	0.0	86.9	13.1	13.1
Albania	med	0.78	0.07	0.0	59.5	40.5	40.5	0.04	0.0	66.8	33.2	33.2
Thailand	med	0.78	0.09	0.1	56.9	43.0	43.1	0.02	0.0	80.0	20.0	20.0
Venezuela	med	0.78	0.09	0.1	57.1	42.9	42.9	0.02	0.0	79.8	20.2	20.2
Colombia	med	0.79	0.09	0.1	54.4	45.6	45.6	0.02	0.0	72.1	27.9	27.9
Brazil	med	0.79	0.08	0.0	53.9	46.1	46.1	0.02	0.0	62.6	37.4	37.4
Mauritius	high	0.80	0.09	0.0	50.0	50.0	50.0	0.01	0.0	50.0	50.0	50.0
Malaysia	high	0.81	0.10	0.1	47.9	52.1	47.9	0.01	0.0	24.4	75.6	24.4
Romania	high	0.81	0.09	0.1	47.8	52.1	47.9	0.05	0.0	46.3	53.7	46.3
Panama	high	0.81	0.08	0.0	45.7	54.3	45.7	0.04	0.0	39.9	60.1	39.9
Trinidad and Tobago	high	0.81	0.11	0.2	46.5	53.4	46.6	0.01	0.0	26.4	73.6	26.4
Oman	high	0.81	0.06	0.0	43.7	56.3	43.7	0.07	0.0	44.2	55.8	44.2
Bulgaria	high	0.82	0.09	0.0	42.8	57.2	42.8	0.03	0.0	29.6	70.4	29.6
Mexico	high	0.82	0.10	0.1	41.8	58.1	41.9	0.01	0.0	3.9	96.1	3.9
United Arab Emirates	high	0.84	0.05	0.0	21.1	78.9	21.1	0.02	0.0	3.4	96.6	3.4
Costa Rica	high	0.84	0.12	0.3	36.5	63.2	36.8	0.01	0.0	0.1	99.9	0.1
Uruguay	high	0.85	0.10	0.0	30.8	69.2	30.8	0.01	0.0	0.0	100.0	0.0
Chile	high	0.86	0.11	0.0	29.1	70.9	29.1	0.01	0.0	0.0	100.0	0.0
Argentina	high	0.86	0.08	0.0	21.5	78.5	21.5	0.01	0.0	0.0	100.0	0.0
Hungary	high	0.87	0.08	0.0	19.7	80.3	19.7	0.02	0.0	0.0	100.0	0.0
Portugal	high	0.90	0.07	0.0	6.7	93.3	6.7	0.03	0.0	0.1	99.9	0.1
Korea, Rep. of	high	0.91	0.07	0.0	6.6	93.4	6.6	0.02	0.0	0.0	100.0	0.0
Greece	high	0.92	0.08	0.0	6.3	93.7	6.3	0.02	0.0	0.0	100.0	0.0
Hong Kong, China	high	0.93	0.05	0.0	0.9	99.1	0.9	0.02	0.0	0.0	100.0	0.0
Israel	high	0.93	0.06	0.0	1.2	98.8	1.2	0.01	0.0	0.0	100.0	0.0
New Zealand	high	0.94	0.06	0.0	1.1	98.9	1.1	0.01	0.0	0.0	100.0	0.0

continued on next page

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Spain	high	0.94	0.06	0.0	1.0	99.0	1.0	0.01	0.0	0.0	100.0	0.0
United Kingdom	high	0.94	0.05	0.0	0.1	99.9	0.1	0.01	0.0	0.0	100.0	0.0
Italy	high	0.94	0.06	0.0	0.9	99.1	0.9	0.01	0.0	0.0	100.0	0.0
France	high	0.94	0.05	0.0	0.3	99.7	0.3	0.01	0.0	0.0	100.0	0.0
Denmark	high	0.94	0.04	0.0	0.0	100.0	0.0	0.01	0.0	0.0	100.0	0.0
Austria	high	0.94	0.05	0.0	0.2	99.8	0.2	0.01	0.0	0.0	100.0	0.0
Belgium	high	0.95	0.04	0.0	0.0	100.0	0.0	0.01	0.0	0.0	100.0	0.0
Switzerland	high	0.95	0.05	0.0	0.1	99.9	0.1	0.01	0.0	0.0	100.0	0.0
Netherlands	high	0.95	0.04	0.0	0.0	100.0	0.0	0.01	0.0	0.0	100.0	0.0
United States	high	0.95	0.03	0.0	0.0	100.0	0.0	0.00	0.0	0.0	100.0	0.0
Japan	high	0.95	0.05	0.0	0.1	99.9	0.1	0.01	0.0	0.0	100.0	0.0
Sweden	high	0.95	0.05	0.0	0.0	100.0	0.0	0.01	0.0	0.0	100.0	0.0
Ireland	high	0.96	0.05	0.0	0.2	99.8	0.2	0.01	0.0	0.0	100.0	0.0
Australia	high	0.96	0.05	0.0	0.0	100.0	0.0	0.02	0.0	0.0	100.0	0.0
Norway	high	0.97	0.04	0.0	0.0	100.0	0.0	0.01	0.0	0.0	100.0	0.0
Expected # of countries misclassified			20.7					10.4				

Table 3.3: Overall error statistics and simulated rank deviations

			Measures based on overall error				
Country i	Expected absolute rank deviation from rank in 2006	2006 HDI	$\sigma_{(overall),i}$	Prob{i='low'}	Prob{i='mid'}	Prob{i='high'}	Prob{i=mis-classified}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Niger	3.7	0.31	0.13	92.7	7.3	0.0	7.3
Mali	4.3	0.34	0.12	91.0	8.9	0.0	9.0
Burkina Faso	4.8	0.34	0.11	91.9	8.1	0.0	8.1
Chad	5.3	0.37	0.11	87.9	12.1	0.0	12.1
Ethiopia	5.7	0.37	0.11	87.2	12.8	0.0	12.8
Burundi	6.2	0.38	0.11	84.9	15.1	0.0	15.1
Mozambique	6.5	0.39	0.12	82.4	17.6	0.0	17.6
Malawi	6.8	0.40	0.13	78.6	21.3	0.1	21.4
Zambia	6.9	0.41	0.10	83.5	16.5	0.0	16.5
Cote d'Ivoire	7.0	0.42	0.09	81.4	18.6	0.0	18.6
Benin	7.1	0.43	0.10	75.7	24.3	0.0	24.3
Tanzania	7.3	0.43	0.09	79.0	21.0	0.0	21.0
Nigeria	7.3	0.45	0.11	67.5	32.4	0.1	32.5
Rwanda	7.1	0.45	0.10	69.7	30.3	0.0	30.3
Senegal	7.1	0.46	0.08	68.2	31.8	0.0	31.8
Mauritania	7.1	0.49	0.09	56.1	43.8	0.0	43.9
Kenya	7.1	0.49	0.08	54.5	45.5	0.0	45.5
Zimbabwe	7.0	0.49	0.06	55.9	44.1	0.0	44.1
Lesotho	7.0	0.49	0.07	53.4	46.6	0.0	46.6
Togo	7.1	0.50	0.09	52.3	47.7	0.0	47.7
Uganda	7.2	0.50	0.10	49.2	50.7	0.1	49.3
Cameroon	7.3	0.51	0.08	47.1	52.8	0.0	47.2
Madagascar	7.4	0.51	0.09	45.8	54.2	0.0	45.8

continued on next page

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sudan	7.7	0.52	0.08	42.0	58.0	0.0	42.0
Congo	7.9	0.52	0.09	41.6	58.2	0.2	41.8
Papua New Guinea	8.1	0.52	0.07	37.0	63.0	0.0	37.0
Nepal	8.5	0.53	0.09	38.3	61.6	0.1	38.4
Bangladesh	8.6	0.53	0.08	35.8	64.2	0.1	35.8
Ghana	9.2	0.53	0.07	33.4	66.5	0.0	33.5
Pakistan	9.5	0.54	0.07	29.6	70.3	0.0	29.7
Lao People's Dem. R.	9.8	0.55	0.10	29.7	69.7	0.6	30.3
Botswana	10.1	0.57	0.06	13.5	86.5	0.0	13.5
India	10.2	0.61	0.06	3.4	96.5	0.1	3.5
Morocco	10.3	0.64	0.04	0.0	99.9	0.0	0.1
South Africa	10.4	0.65	0.09	3.9	91.6	4.5	8.4
Guatemala	10.5	0.67	0.05	0.0	99.7	0.3	0.3
Honduras	10.8	0.68	0.06	0.1	97.1	2.8	2.9
Mongolia	10.7	0.69	0.09	2.0	85.9	12.1	14.1
Bolivia	10.6	0.69	0.06	0.1	96.0	3.9	4.0
Nicaragua	10.7	0.70	0.06	0.1	94.3	5.6	5.7
Egypt	10.6	0.70	0.05	0.0	96.9	3.1	3.1
Viet Nam	10.5	0.71	0.08	0.3	88.2	11.5	11.8
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Indonesia	10.4	0.71	0.07	0.1	91.1	8.8	8.9
Syrian Arab Republic	10.5	0.72	0.09	1.0	80.7	18.3	19.3
Jamaica	10.4	0.72	0.07	0.1	86.5	13.4	13.5
Algeria	10.2	0.73	0.06	0.0	90.5	9.5	9.5
El Salvador	10.2	0.73	0.07	0.1	82.9	17.0	17.1
Iran, Islamic Rep. of	10.2	0.75	0.06	0.0	81.6	18.4	18.4
Dominican Republic	10.1	0.75	0.06	0.0	79.1	20.9	20.9
Sri Lanka	10.0	0.76	0.08	0.1	71.3	28.6	28.7
Paraguay	9.8	0.76	0.08	0.0	71.2	28.8	28.8
Turkey	9.8	0.76	0.08	0.0	71.6	28.4	28.4
Jordan	9.6	0.76	0.07	0.0	70.6	29.3	29.4
Tunisia	9.5	0.76	0.06	0.0	74.6	25.4	25.4

continued on next page

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Philippines	9.7	0.76	0.06	0.0	72.0	27.9	28.0
Peru	9.5	0.77	0.05	0.0	73.3	26.7	26.7
China	9.6	0.77	0.06	0.0	69.2	30.8	30.8
Lebanon	9.8	0.77	0.07	0.0	65.2	34.8	34.8
Saudi Arabia	9.8	0.78	0.07	0.0	62.6	37.4	37.4
Albania	9.9	0.78	0.08	0.0	58.3	41.7	41.7
Thailand	10.0	0.78	0.09	0.1	56.7	43.2	43.3
Venezuela	10.0	0.78	0.09	0.1	56.8	43.1	43.2
Colombia	10.3	0.79	0.09	0.1	54.2	45.7	45.8
Brazil	10.4	0.79	0.09	0.0	53.7	46.3	46.3
Mauritius	10.5	0.80	0.09	0.0	50.0	50.0	50.0
Malaysia	10.5	0.81	0.10	0.1	47.9	52.1	47.9
Romania	10.7	0.81	0.11	0.2	47.9	51.9	48.1
Panama	10.7	0.81	0.09	0.0	46.0	54.0	46.0
Trinidad and Tobago	10.7	0.81	0.11	0.2	46.5	53.3	46.7
Oman	10.8	0.81	0.09	0.0	45.6	54.3	45.7
Bulgaria	11.0	0.82	0.09	0.0	43.1	56.8	43.2
Mexico	11.0	0.82	0.10	0.1	41.9	58.0	42.0
United Arab Emirates	10.9	0.84	0.05	0.0	23.1	76.9	23.1
Costa Rica	10.9	0.84	0.12	0.3	36.6	63.1	36.9
Uruguay	10.9	0.85	0.10	0.0	30.8	69.1	30.9
Chile	10.8	0.86	0.11	0.0	29.2	70.8	29.2
Argentina	10.6	0.86	0.08	0.0	21.8	78.2	21.8
Hungary	10.5	0.87	0.08	0.0	20.1	79.9	20.1
Portugal	10.2	0.90	0.08	0.0	8.7	91.3	8.7
Korea, Rep. of	9.9	0.91	0.08	0.0	7.0	93.0	7.0
Greece	9.6	0.92	0.08	0.0	6.6	93.4	6.6
Hong Kong, China	9.2	0.93	0.06	0.0	1.3	98.7	1.3
Israel	9.0	0.93	0.06	0.0	1.4	98.6	1.4
New Zealand	8.6	0.94	0.06	0.0	1.2	98.8	1.2

continued on next page

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Spain	8.3	0.94	0.06	0.0	1.2	98.8	1.2
United Kingdom	7.9	0.94	0.05	0.0	0.2	99.8	0.2
Italy	7.7	0.94	0.06	0.0	1.0	99.0	1.0
France	7.4	0.94	0.05	0.0	0.3	99.7	0.3
Denmark	7.2	0.94	0.04	0.0	0.0	100.0	0.0
Austria	7.2	0.94	0.05	0.0	0.2	99.8	0.2
Belgium	7.2	0.95	0.05	0.0	0.1	99.9	0.1
Switzerland	7.2	0.95	0.05	0.0	0.1	99.9	0.1
Netherlands	7.4	0.95	0.04	0.0	0.1	99.9	0.1
United States	7.8	0.95	0.03	0.0	0.0	100.0	0.0
Japan	8.3	0.95	0.05	0.0	0.1	99.9	0.1
Sweden	9.0	0.95	0.05	0.0	0.1	99.9	0.1
Ireland	9.7	0.96	0.05	0.0	0.2	99.8	0.2
Australia	11.3	0.96	0.05	0.0	0.1	99.9	0.1
Norway	14.0	0.97	0.04	0.0	0.0	100.0	0.0
Average world absolute deviation in rank	9.0		Expected # of countries misclassified.				22.9

Table 3.4: As of 2006, countries misclassified due to the arbitrary cut off points

Countries with $\mathbf{HDI}_{2006} \in$ [0.5 and 0.55)	**Countries with $\mathbf{HDI}_{2006} \in$ [0.7 and 0.8)**
Bangladesh	Albania
Cameroon	Brazil
Congo	China
Ghana	Colombia
Madagascar	Dominican Republic
Nepal	Algeria
Pakistan	Egypt
Papua New Guinea	Indonesia
Sudan	Iran, Islamic Rep. of
Uganda	Jamaica
	Jordan
	Lebanon
	Sri Lanka
	Peru
	Philippines
	Paraguay
	Saudi Arabia
	El Salvador
	Syrian Arab Republic
	Thailand
	Tunisia
	Turkey
	Venezuela
	Vietnam

Table 3.5: Updating error summary statistics for the period 1999 to 2005

Indicators	Non-industrialized Countries				Industrialized Countries				Industrial vs. Non-industrialized Countries	
	mean	std. dev.	min	max	mean	std. dev.	min	max	Difference in means	Ratio of std. dev
HDI	0.01	0.02	-0.06	0.08	0.01	0.01	-0.03	0.05	0.002^{*}	0.55
Health	0.00	0.03	-0.14	0.11	0.00	0.01	-0.01	0.02	$0.002^{\dagger}$	0.20
Education	0.00	0.03	-0.11	0.25	0.00	0.01	-0.09	0.05	-0.004^{*}	0.44
Income	0.01	0.02	-0.07	0.15	0.02	0.02	-0.02	0.09	0.009^{*}	0.95

Symbol * states that estimate of 'Differences in means' is statically significant at 1% level, tested by regressing the vector of updating errors on a constant and an indicator variable that takes the value one if the country is industrialized and zero otherwise using robust standard errors. Symbol † indicates that the estimate of the same regression is different from zero only at the 15% significance level.

Table 3.6: Upgrading error statistics of variables underlying the HDI

country	HDI 2006	GDP per capita 2006 [PPP US$]	$\sigma(\mathrm{GDP})_i$	Gross enroll-ment ratio (GER) 2006 [%]	$\sigma(\mathrm{GER})_i$	Adult liter-acy rate (ALR) 2006 [%]	$\sigma(\mathrm{ALR})_i$	Life ex-pectancy (LE) at birth 2006 [years]	$\sigma(\mathrm{LE})_i$
Niger	0.311	779.1	67.7	21.5	0.6	18.2	1.2	44.6	2.4
Mali	0.338	997.8	55.3	35.0	1.1	28.7	11.1	48.1	2.6
Burkina Faso	0.342	1168.8	75.5	26.4	1.0	27.6	6.5	47.9	0.7
Chad	0.368	2090.1	73.5	34.8	1.5	49.1	10.2	43.7	1.0
Ethiopia	0.371	755.8	86.3	36.0	1.1	44.0	0.5	47.8	1.4
Burundi	0.384	677.3	49.2	36.2	5.1	52.8	2.7	44.0	1.4
Mozambique	0.390	1236.6	59.2	48.6	4.9	49.0	0.6	41.6	2.2
Malawi	0.400	646.2	59.6	64.3	1.7	63.5	0.6	39.8	1.0
Zambia	0.407	943.2	69.4	54.3	2.9	81.5	4.8	37.7	3.4
Cote d'Ivoire	0.421	1551.0	90.3	39.6	1.2	52.7	1.2	45.9	2.5
Benin	0.428	1091.0	136.4	49.4	1.4	42.0	3.3	54.3	1.4
Tanzania	0.430	674.4	36.7	47.8	4.1	79.1	3.4	45.9	2.8
Nigeria	0.448	1154.2	56.1	55.0	5.7	69.5	0.5	43.4	2.9
Rwanda	0.450	1262.7	254.9	51.8	6.6	71.6	2.4	44.2	3.9
Senegal	0.460	1712.8	144.6	38.1	0.9	41.2	0.3	56.0	1.0
Mauritania	0.486	1940.5	277.3	45.6	1.6	42.2	3.5	53.1	1.5
Kenya	0.491	1139.6	73.8	60.1	2.6	86.0	4.4	47.5	1.8
Zimbabwe	0.491	2065.2	352.4	52.4	2.6	91.3	1.8	36.6	2.5
Lesotho	0.494	2618.9	247.9	65.5	2.9	85.3	1.6	35.2	3.7
Togo	0.495	1535.8	178.6	55.0	3.3	62.2	3.1	54.5	2.3
Uganda	0.502	1478.4	93.4	66.1	10.7	70.7	0.4	48.4	1.2
Cameroon	0.506	2173.6	194.5	62.3	2.7	75.8	0.3	45.7	2.3
Madagascar	0.509	857.0	57.2	56.5	2.2	69.7	6.5	55.6	2.7
Sudan	0.516	1948.7	400.5	36.7	0.9	62.1	0.9	56.5	0.3
Congo	0.520	978.2	279.0	51.7	11.4	84.9	0.4	52.3	1.5

continued on next page

country	HDI 2006	GDP per capita 2006 [PPP US$]	$\sigma(\text{GDP})_i$	Gross enroll-ment ratio (GER) 2006 [%]	$\sigma(\text{GER})_i$	Adult liter-acy rate (ALR) 2006 [%]	$\sigma(\text{ALR})_i$	Life expectancy (LE) at birth 2006 [years]	$\sigma(\text{LE})_i$
Papua N. G.	0.523	2543.4	146.3	40.7	1.6	66.7	6.0	55.7	1.5
Nepal	0.527	1489.8	75.2	57.0	2.3	46.3	1.3	62.1	0.7
Bangladesh	0.530	1870.3	163.4	57.1	7.7	42.1	0.8	63.3	0.6
Ghana	0.532	2239.7	159.4	47.2	2.3	76.0	7.8	57.0	1.5
Pakistan	0.539	2225.4	118.1	38.4	3.2	46.6	2.2	63.4	2.2
Lao	0.553	1953.9	122.7	61.0	0.5	68.1	8.7	55.1	0.4
Botswana	0.570	9944.7	728.2	70.7	3.7	80.5	0.4	34.9	4.0
India	0.611	3139.4	258.1	62.0	1.3	60.3	1.2	63.6	0.4
Morocco	0.640	4309.4	134.1	57.8	1.7	52.6	0.4	70.0	0.4
South Africa	0.653	11192.2	1017.3	76.6	6.6	86.7	1.5	47.0	1.6
Guatemala	0.673	4313.0	304.6	66.2	2.8	71.2	0.6	67.6	0.5
Honduras	0.683	2876.4	216.5	71.4	1.7	77.3	2.2	68.1	1.7
Mongolia	0.691	2055.6	131.8	77.3	2.0	98.6	13.6	64.5	1.9
Bolivia	0.692	2719.6	221.5	86.5	5.7	87.7	0.3	64.4	0.3
Nicaragua	0.698	3634.2	361.5	70.2	2.1	67.8	4.6	70.0	0.2
Egypt	0.702	4210.8	139.5	75.5	1.8	58.5	0.9	70.2	0.5
Viet Nam	0.709	2744.8	84.0	62.8	1.3	93.2	1.6	70.8	0.5
Indonesia	0.711	3608.5	175.0	68.4	2.0	89.0	0.2	67.2	0.2
Syria	0.716	3609.8	542.5	62.6	3.1	77.7	3.1	73.6	0.9
Jamaica	0.724	4163.1	166.0	76.9	5.0	88.3	0.1	70.7	1.6
Algeria	0.728	6603.1	449.3	73.2	1.8	71.0	1.3	71.4	0.5
El Salvador	0.729	5040.8	688.1	69.7	1.5	80.6	0.2	71.1	0.2
Iran	0.746	7524.8	287.1	72.2	2.9	80.1	1.1	70.7	0.6
Dominican R.	0.751	7449.3	461.8	74.1	1.7	85.1	1.1	67.5	2.0
Sri Lanka	0.755	4389.6	214.3	62.7	3.0	92.5	0.7	74.3	0.8
Paraguay	0.757	4812.9	552.4	69.7	3.7	94.2	1.1	71.2	0.1
Turkey	0.757	7752.6	209.6	69.1	2.5	87.0	0.6	68.9	0.6

continued on next page

country	HDI 2006	GDP per capita 2006 [PPP US$]	$\sigma(\text{GDP})_i$	Gross enrollment ratio (GER) 2006 [%]	$\sigma(\text{GER})_i$	Adult literacy rate (ALR) 2006 [%]	$\sigma(\text{ALR})_i$	Life expectancy (LE) at birth 2006 [years]	$\sigma(\text{LE})_i$
Jordan	0.760	4687.8	320.0	79.0	11.0	91.9	0.6	71.6	0.3
Tunisia	0.760	7767.6	227.5	75.4	0.8	75.2	0.1	73.5	1.2
Philippines	0.763	4614.1	217.9	81.5	1.1	95.8	1.6	70.7	0.1
Peru	0.767	5678.4	218.7	86.4	2.9	91.2	2.3	70.2	0.2
China	0.768	5896.1	127.2	70.4	2.5	87.6	2.4	71.9	0.3
Lebanon	0.774	5836.8	653.7	83.8	1.9	87.8	0.4	72.2	1.2
Saudi Arabia	0.777	13825.2	1479.4	58.6	1.6	79.6	0.4	72.0	0.4
Albania	0.784	4977.8	426.7	68.0	0.8	87.2	5.9	73.9	0.1
Thailand	0.784	8089.8	303.7	73.7	4.7	96.2	1.7	70.3	0.6
Venezuela	0.784	6042.7	991.1	74.2	1.8	93.7	0.3	73.0	0.3
Colombia	0.790	7256.3	415.2	72.9	2.3	92.7	0.6	72.6	0.2
Brazil	0.792	8194.7	439.7	85.7	6.4	88.5	0.8	70.8	0.8
Mauritius	0.800	12027.3	506.4	74.5	1.5	86.0	0.6	72.4	0.4
Malaysia	0.805	10276.1	331.0	73.2	2.2	89.4	0.2	73.4	0.1
Romania	0.805	8479.5	756.9	75.3	2.1	98.5	0.5	71.5	0.4
Panama	0.809	7277.8	756.8	79.7	2.1	92.8	0.3	75.0	0.1
Trinidad & T.	0.809	12181.9	339.7	66.9	1.9	98.7	2.4	69.8	1.4
Oman	0.810	15259.1	1451.6	68.3	2.1	77.1	0.6	74.3	0.8
Bulgaria	0.816	8077.9	646.3	80.9	2.5	98.7	0.2	72.4	0.5
Mexico	0.821	9803.2	406.7	75.3	0.7	92.3	0.8	75.3	0.6
Arab. Emirat.	0.839	24055.9	1859.6	59.9	4.7	78.3	0.3	78.3	1.1
Costa Rica	0.841	9481.4	1016.6	72.4	1.2	96.1	0.1	78.3	0.6
Uruguay	0.851	9420.6	460.2	89.4	1.3	97.9	0.1	75.6	0.2
Chile	0.859	10873.6	1458.5	81.3	1.0	96.3	0.2	78.1	0.6
Argentina	0.863	13298.0	1038.5	89.3	4.8	97.2	0.1	74.6	0.2
Hungary	0.869	16814.4	1232.3	87.5	2.0	99.0	0.1	73.0	0.4

continued on next page

country	HDI 2006	GDP per capita 2006 [PPP US$]	$\sigma(\text{GDP})_i$	Gross enrollment ratio (GER) 2006 [%]	$\sigma(\text{GER})_i$	Adult literacy rate (ALR) 2006 [%]	$\sigma(\text{ALR})_i$	Life expectancy (LE) at birth 2006 [years]	$\sigma(\text{LE})_i$
Portugal	0.904	19628.9	970.5	89.3	2.5	99.0	0.7	77.5	0.3
Korea, R.	0.912	20499.3	1002.1	95.0	2.0	99.0	0.3	77.3	0.8
Greece	0.921	22204.7	908.5	93.4	1.8	99.0	2.3	78.3	0.2
Hong Kong	0.927	30822.1	967.0	76.7	4.5	94.3	0.3	81.8	0.5
Israel	0.927	24381.6	1066.0	89.7	2.6	95.8	0.8	80.0	0.2
New Zealand	0.936	23413.0	1027.2	100.0	2.4	99.0	0.0	79.3	0.3
Spain	0.938	25046.8	1268.7	96.1	1.9	99.0	0.2	79.7	0.2
Italy	0.940	28180.2	1831.0	89.3	3.1	99.0	0.1	80.2	0.4
UK	0.940	30821.2	1427.8	93.1	7.5	99.0	0.0	78.5	0.0
France	0.942	29300.5	1421.8	92.6	1.2	99.0	0.0	79.6	0.2
Denmark	0.943	31913.8	2203.3	100.0	2.3	99.0	0.0	77.3	0.1
Austria	0.944	32276.4	1840.6	91.1	1.7	99.0	0.0	79.2	0.2
Belgium	0.945	31095.8	1158.4	94.7	4.4	99.0	0.0	79.1	0.3
Netherlands	0.947	31789.4	2108.1	98.2	2.1	99.0	0.0	78.5	0.0
Switzerland	0.947	33039.6	1606.2	85.7	2.8	99.0	0.0	80.7	0.4
United States	0.948	39676.1	1870.0	93.3	1.3	99.0	0.0	77.5	0.2
Japan	0.949	29251.4	1534.0	85.5	1.9	99.0	0.0	82.2	0.2
Sweden	0.951	29540.7	1269.6	96.5	6.5	99.0	0.0	80.3	0.3
Ireland	0.956	38827.0	1826.0	99.0	1.8	99.0	0.0	77.9	0.3
Australia	0.957	30331.1	3492.0	100.0	5.7	99.0	0.0	80.5	0.3
Norway	0.965	38453.5	4225.7	100.0	0.9	99.0	0.0	79.6	0.1

Table 3.7: Official development assistance (ODA) received in US dollar per capita by year and human development category

	2006	2005	2004	2003	2002	2001
'medium'	7.2	6.5	6.5	5.7	5.9	6.6
'low'	30.1	27.9	24.2	18.4	14.9	14.5

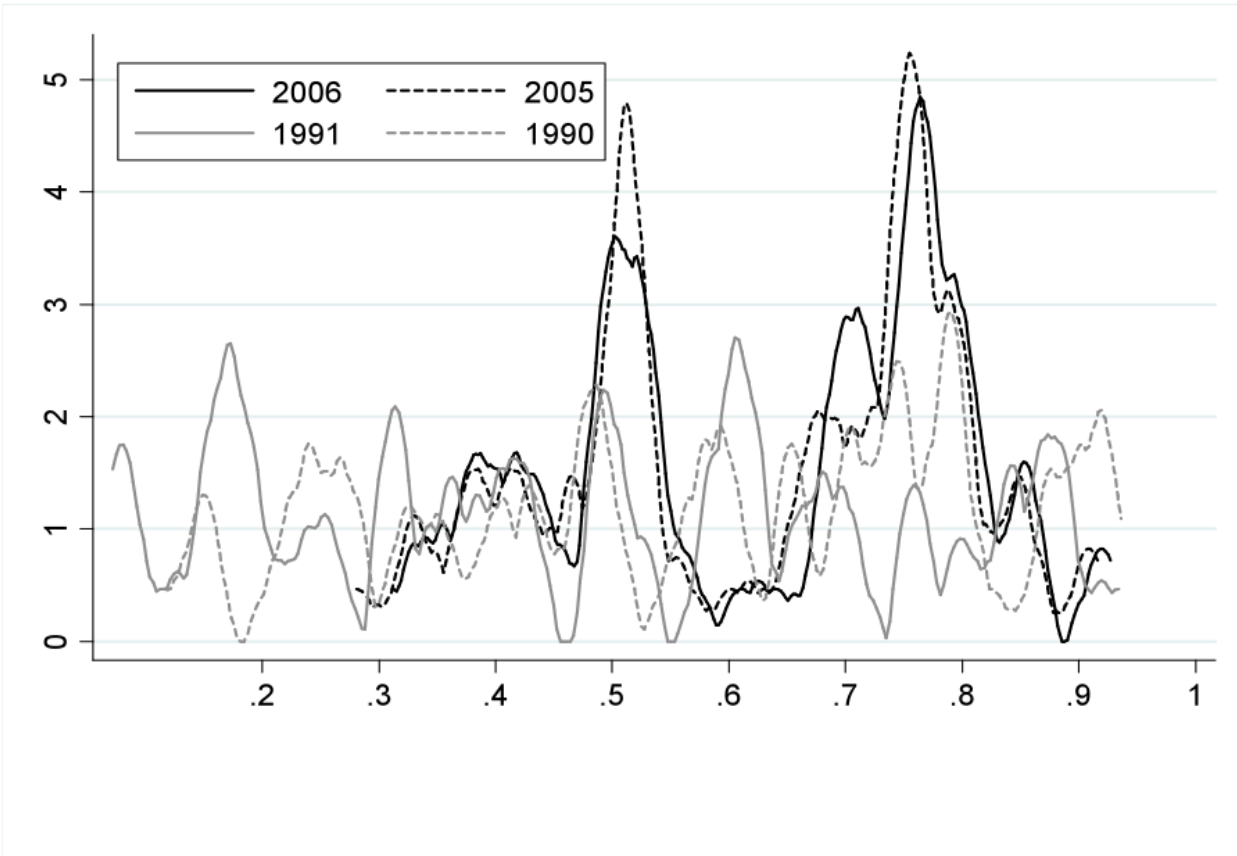
Data are from the Human Development Reports 2001 to 2006.

Table 3.8: Convergence club regression results for medium development category

Sample conditional on	$HDI_{2006} \in [0.5,0.8)$	$HDI_{2006} \in [0.55,0.70)$
constant α	-.02556 (-56.69)	-.02847 (-35.36)
slope β	-.01380 (- 6.74)	-.02667 (-4.59)
adjusted R^2	.53	.74

t statistics in parentheses.

Figure 3.1: Historical HDI scores for Non-industrialized Countries in 1990/91 and 2005/06



Notes: On the horizontal axis we display the HDI, which ranges from 0 to 1. 1990/91 are the first and 2005/06 are last two years for which the HDI scores originally have been made available (HDR, 1990, 1991, 2005, 2006). To make the HDI-distributions comparable across years we use the balanced panel of 99 developing countries that have been evaluated by the UNDP for all years. Countries that existed for a subset of years only (e.g. Croatia) are not considered. All densities are estimated by the Epanechnikov kernel method with bandwidth 0.01.

Figure 3.2: HDI of 1975 of Portugal and Venezuela as reported in the years 1999 to 2006

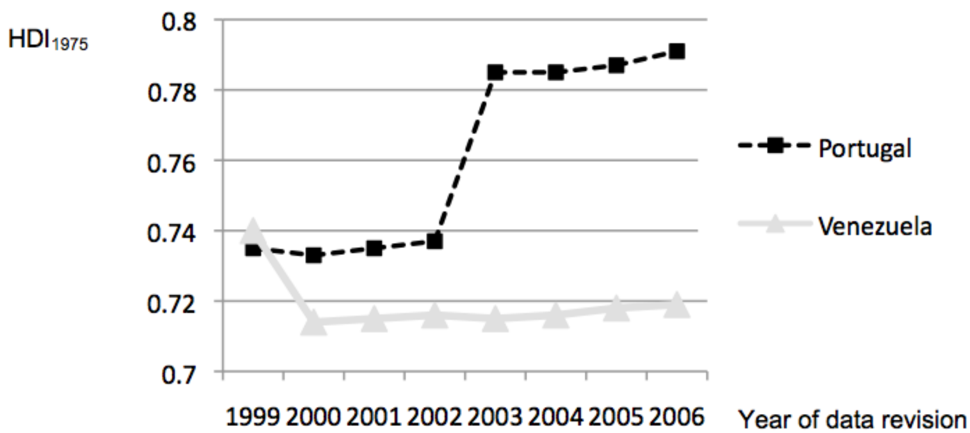


Figure 3.3: Density of HDI as published in the Human Development Reports (HDR)

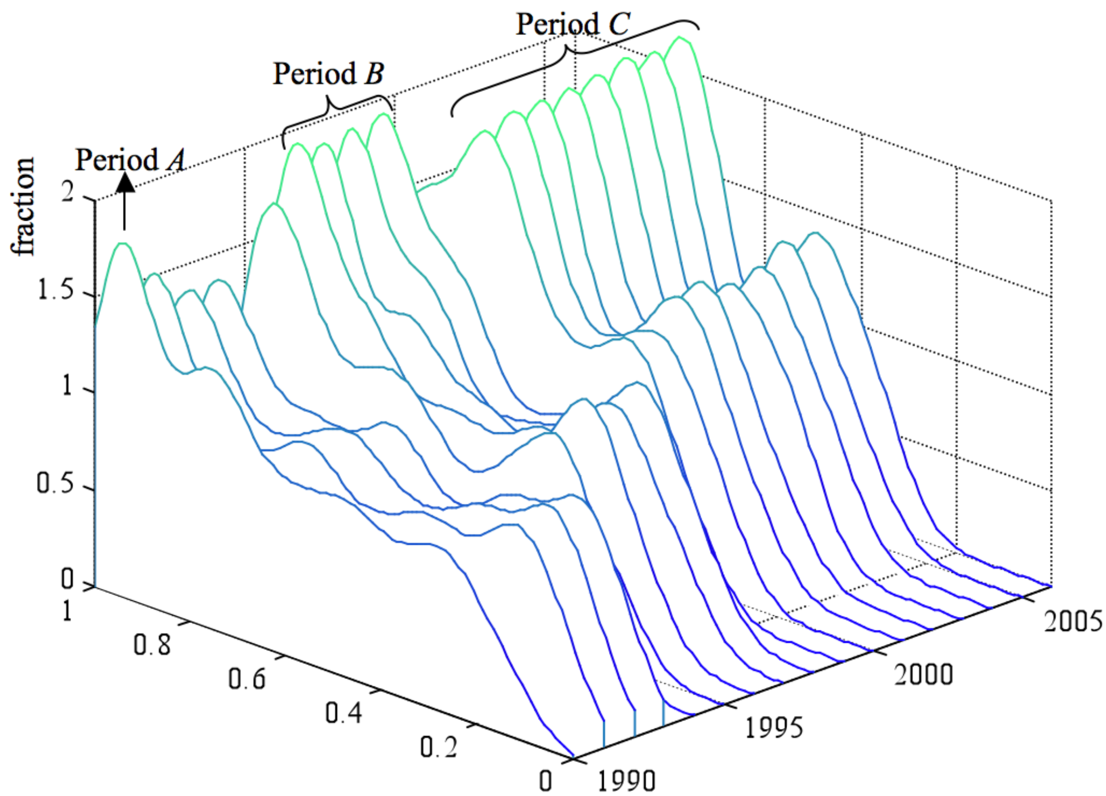
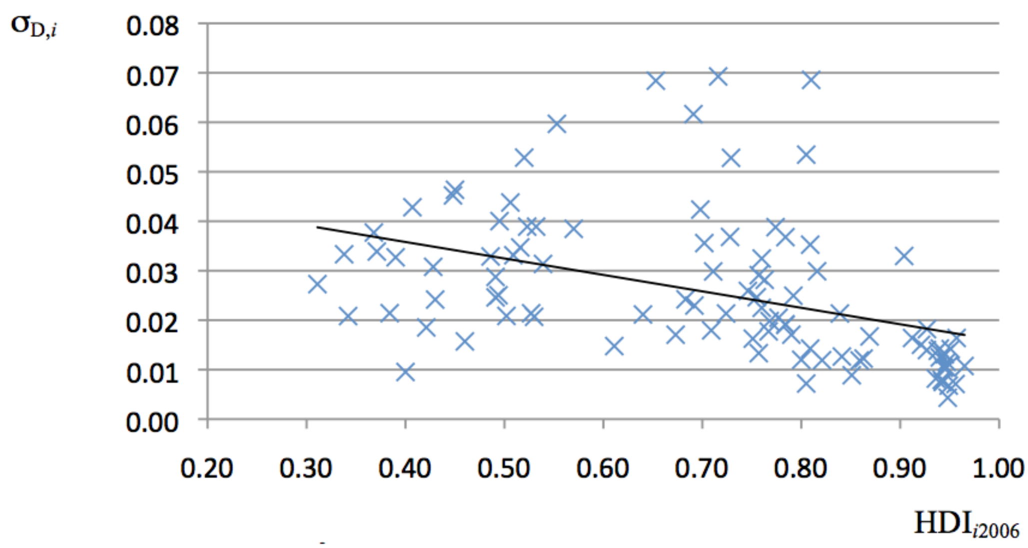


Figure 3.4: Relationship between countries' development status and the standard deviations due to measurement error generated by data updates



Linear trendline based on sample of 99 countries, $R^2 = 0.184$.

Figure 3.5: Representation of data error of a country with HDI = 0.65

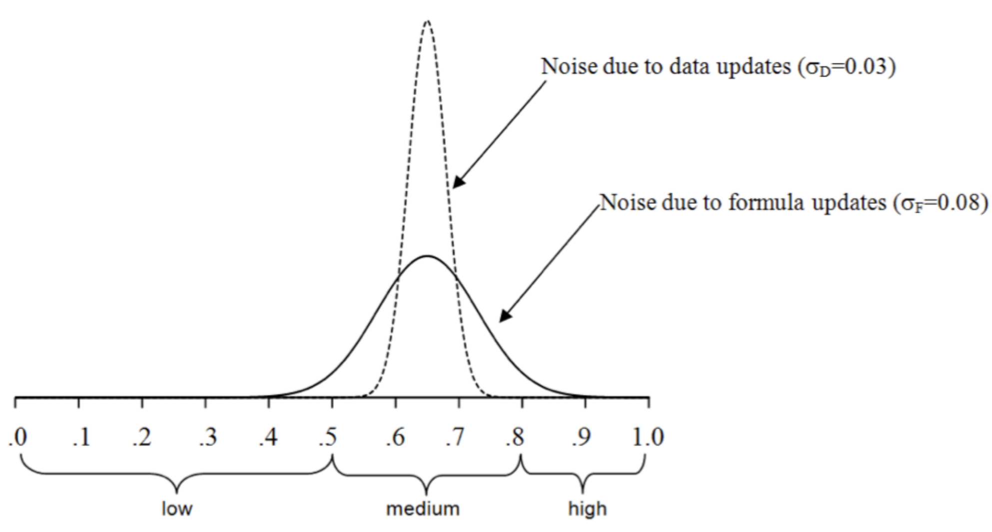
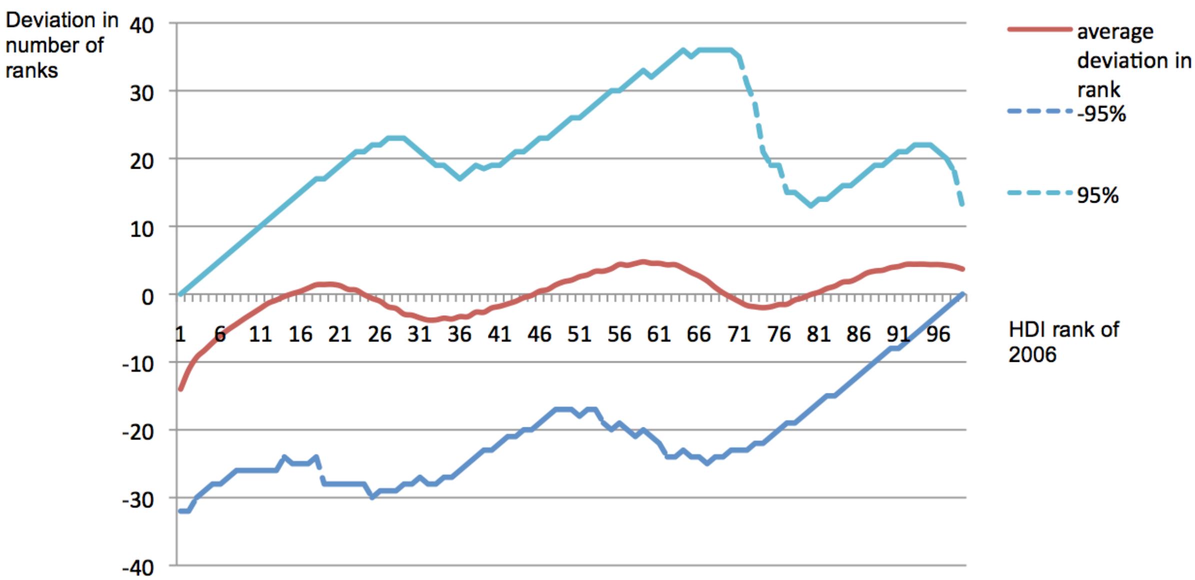
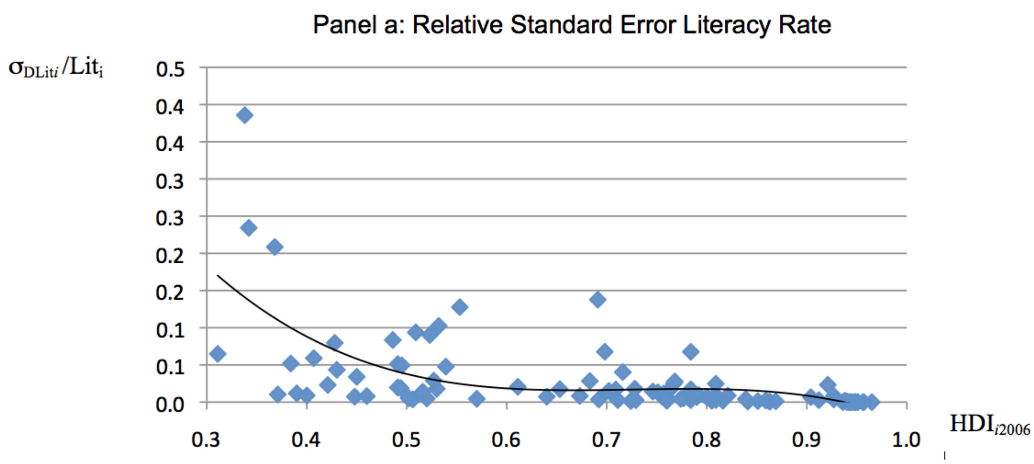


Figure 3.6: Simulated HDI ranks compared to rank of country in 2006

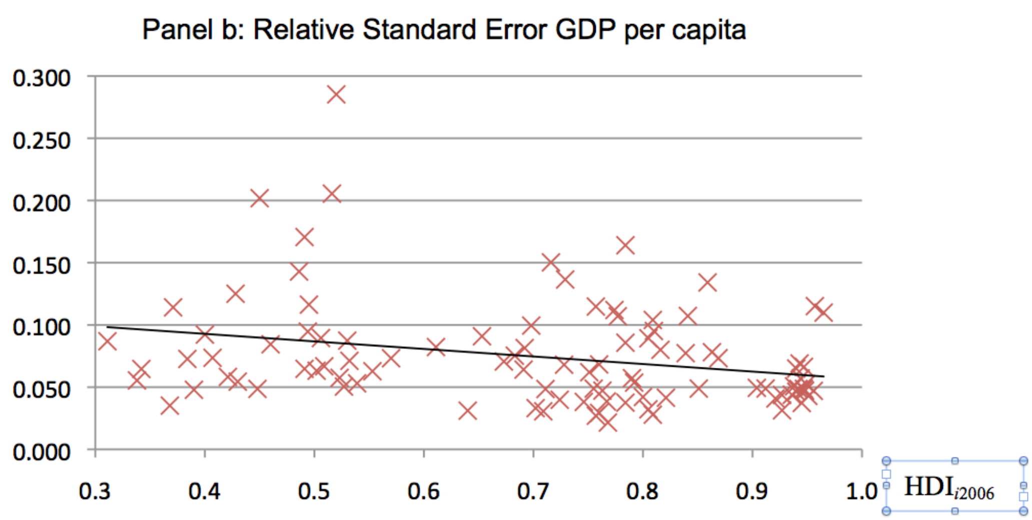


Average, 2.5% and 97.5% percentiles of simulated rank distributions, displaying the deviation in rank for a country compared to its rank in 2006. Ranks based on the sample of 99 countries for which the overall cardinal error can be calculated consistently.

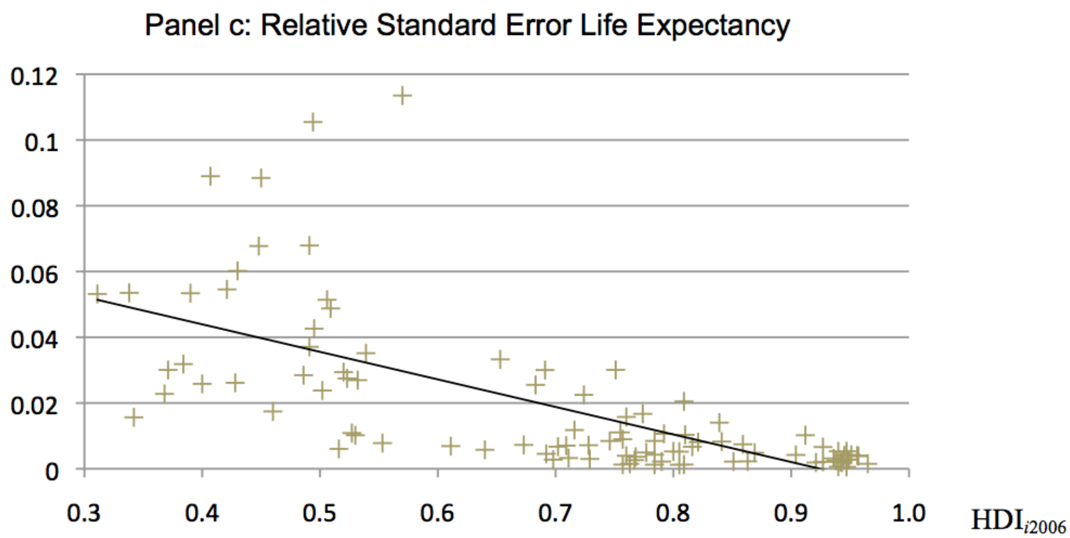
Figure 3.7: Relationship between Countries' Development Status and the relative standard Errors due to Measurement Error generated by Data Updates of the underlying Variables of the HDI



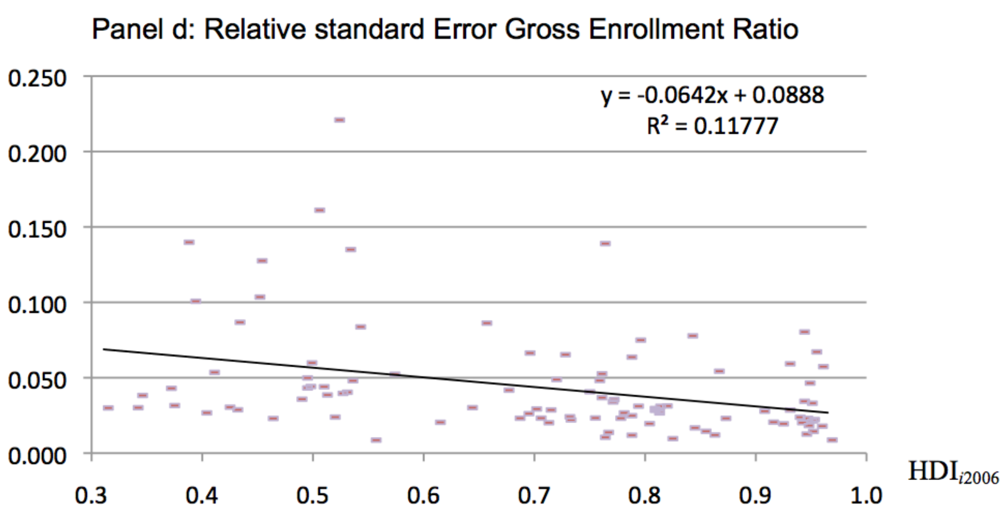
Quadratic trendline y = $-2.2904x^3 + 4.9837x^2 - 3.5914x + 0.8739$ is based on least squares estimation of sample of 99 countries, $R^2 = 0.375$.



Linear trendline is based on least squares estimation of sample of 99 countries, $R^2 = 0.073$.



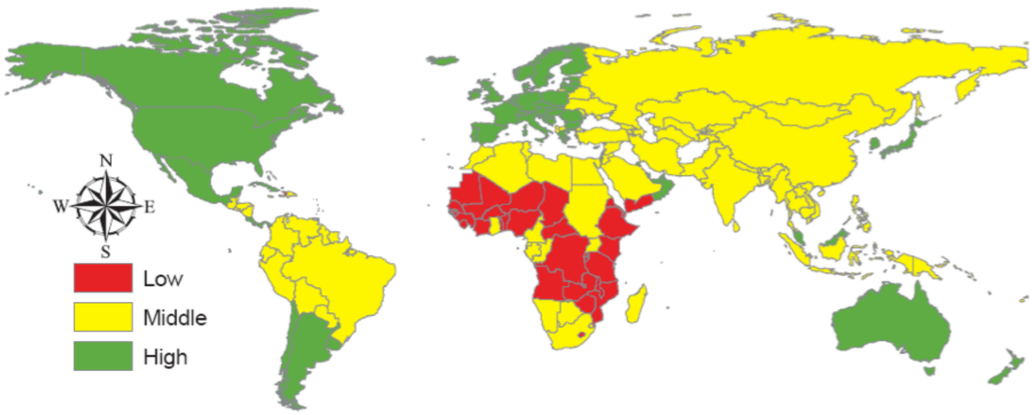
Linear trendline is based on least squares estimation of sample of 99 countries, $R^2 = 0.456$.



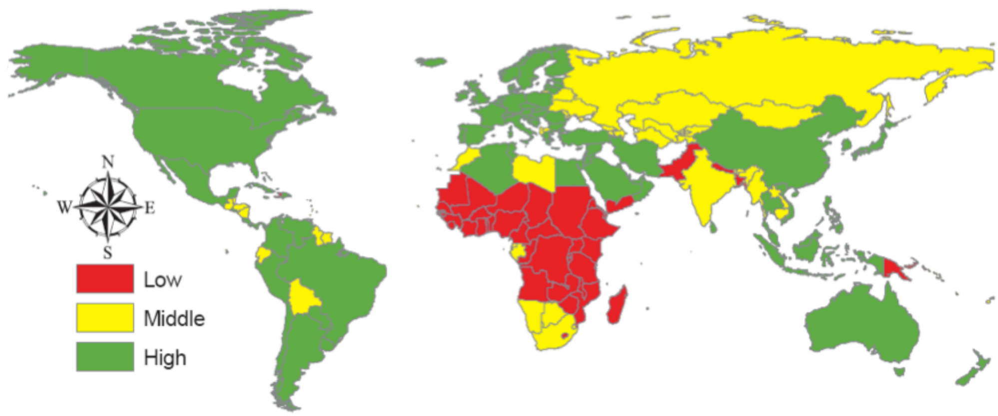
Linear trendline is based on least squares estimation of sample of 99 countries, $R^2 = 0.118$.

Figure 3.8: World map of the Human Development Index

Panel (a). Reported Human Development Index 2006

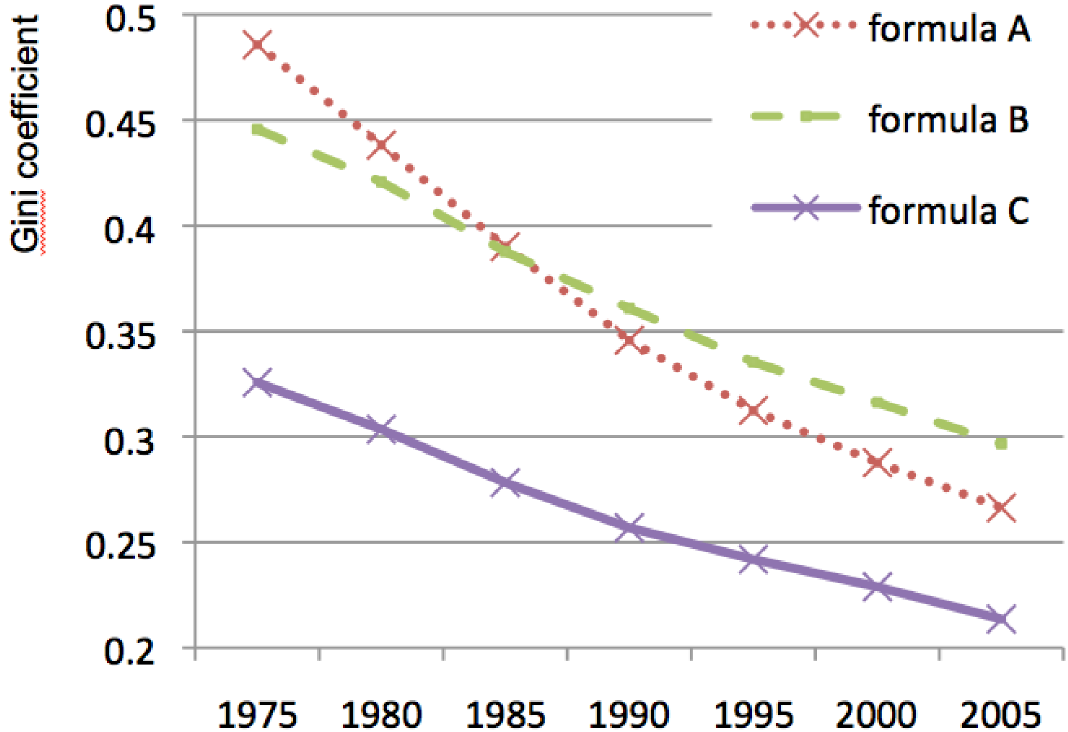


Panel (b). Adjusted Human Development Index 2006



Note: Panel (a) displays the classification using the actually reported HDI Index for the year 2006 for all reported countries (industrialized and non-industrialized). Countries in white have no reported data. Panel (b) displays the classification based on the revised thresholds that we calculate in Section 4.3. if the UNDP had consistently updated the cutoff values for classification.

Figure 3.9: Gini Coefficients computed by the HDI Formulas A, B and C



3.8 Description of formula changes and formula equations

Formula Changes

The most important changes are described in the following subsections, and the reader is referred to Anand and Sen (1994, 1997, 1998), the technical appendices of the HDRs (1990 to 2006) and to Jahan (2000) for details.

Income. At first, in 1990 GDP per capita (in PPP) was logged in the income index. Between 1991 and 1998, however income above a certain cut-off point got substantially adjusted with a regressive version of the "Atkinson Function". The cut-off point was taken as the average world income on the assumption that every person should have at least this level of income for building basic capabilities (Jahan 2000). Since this formulation however was argued to particularly punish middle income countries, the original formulation of logging GDP per capita was again introduced in 1999 with formula h_C (Anand and Sen, 1998).

Education. Following the suggestion by Kelley (1991), compared to 1990, in 1991 mean years of schooling was added as a second component to adult literacy to form a more general index of education. Adult literacy was given two-thirds weight and mean years of schooling one-third weight according to an argument that adult literacy is a more representative stock variable for educational attainment. However, the variable "mean years of schooling" was constructed in a complicated way and for some of the countries it was criticized to not reflect their educational infrastructure properly (Jahan 2000). In an effort to further improve the measure, from 1995 onward, mean year of schooling was replaced by the combined gross enrolment of schooling at the primary, secondary and tertiary level of education.

Maxima and mimina. Until 1994, in order to normalize the variables $\boldsymbol{x}$ into the double bounded indices $\boldsymbol{y}$, observed maxima and minima were used as goalposts. This created however the problem that it got difficult to distinguish whether changes in the HDI of a country was because of its improved performance or because of changes to the maxima and minima of the sample of countries considered. In order to make HDI trends over time more meaningful, since 1994 fixed maxima and minima were introduced based on the trends of the variables of what that their values was estimated to be in the following 25 years (HDR, 1994).

<u>HDI formulas h_f used by the UNDP in the three subperiods $f \in \{A, B, C\}$</u>

Formula h_C

y_{1i} = *health*$_i$ = (*life expectancy*$_i$ – 25) / (85 – 25)

y_{2i} = *education*$_i$ = 2/3 *adult literacy index*$_i$ + 1/3 *combined gross enrollment index*$_i$

y_{3i} = *income*$_i$ = min(1, (log(*GDP per capita*$_i$) – log(100)) / (log(40000) – log(100)))

adult literacy index$_i$ = (*adult literacy rate*$_i$ – 0) / (100 – 0)

gross enrollment index$_i$ =min(1, (*Combined gross enrolment ratio for primary, secondary and tertiary level schools*$_i$ / 100)[19]

Formula h_B

income is calculated by a version of the "Atkinson Function". Given a cutoff value c

W(y*) = y* for y* < c

W(y*) = c + 2(y* – c)ˆ(1/2) for y* in [c, 2c]

W(y*) = c + 2(c)ˆ(1/2) + 3(y* – c)ˆ(1/3) for y* in [2c, 3c]

for [3c, 4c], [4c, 5c] etc.

whereby c is defined as 'world average income'.[20] With this function defined,

y_{3i} = *income*$_i$ = W(y*$_i$) = (W(*GDP per capita*$_i$) – W(100)) / W(40000)

y_{1i} and y_{2i} are defined as in formula h_C

Formula h_A

y_{1i} = *health*$_i$ = (*life expectancy*$_i$ – low$_1$) / (high$_1$ – low$_1$)

y_{2i} = *education*$_i$ = *adult literacy rate*$_i$ – low$_2$) / (high$_2$ – low$_2$)

y_{3i} = *income*$_i$ = log(*GDP per capita*$_i$) – log(low$_{3)}$) / (log(high$_3$) – log(low$_3$))

The "goalposts" of low$_k$ and high$_k$ for each subindicator $k \in$ {1,2,3} are the values of the minimum and the maximum of the kth subindicator index of all considered countries. In 1990, the implicit goalposts were 42 and 78 for life expectancy, 12.3 and 100 for literacy rate, and 10ˆ2.34 and 10ˆ3.68 for GDP per capita.

[19] The combined gross enrolment ratio is calculated for the number of students enrolled in primary, secondary and tertiary levels of education, regardless of age, as percentage of the population of official school age for the three levels. The gross enrolment ratio can be greater than 100% as a result of grade repetition and entry at ages younger or older than the typical age at that grade level. For this reason the education index takes the minimum of one and the ratio.

[20] This function is described on page 111 in the Technical Notes of the HDR of 1995.

Chapter 4

Appendices

4.1 Appendices to Chapter1

4.1.1 Total electricity vs temperature response

This appendix provides estimates that compare the differences in the total electricity use across households of different vintages. Note that because the variation in age of housing does not vary over time, this precludes the use of household fixed effects. Vintage effects will include differences across households that are not related to the building, such as increases in the amount of appliances or televisions.

Regression results are shown below using a random effects specification with clustering at the ZIP9 level in Table 1.5. The first column shows that newer buildings have larger utility bills, with no clear pattern across decades. The second column two adds a control variable for square footage. Size increases total electricity use, as expected, but the estimates have 1990s and 1980s buildings using less energy after controlling for size, whereas 1970s buildings use slightly more than pre1970s buildings. The third column adds controls for temperature interacted with all variables; the signs of the vintage coefficients are unchanged.

Though interesting empirical regularities, the coefficients on the vintage variables are hard to interpret. They can be rationalized both by increasing efficiency of appliances in new buildings or fewer appliances in new buildings of comparable size.

It is important to note that newer buildings have a larger temperature invariant component (Column T1), which means that the same percentage increase in new buildings and old buildings (due to temperature difference) also means a higher change in kWh for the new buildings.

4.1.2 Functional Form

The function form used in electricity regressions varies across studies, with the literature split between have $ln(kWh_{useperday})$ (dubbed ”ln”) or $kWh_{useperday}$ (dubbed ”levels”) as the LHS variable. In many cases, the choice is ad hoc, justified on the grounds that the ln specification compares percent changes across observations which roughly controls for size. In KEMA-XENERGY (2004), a conditional demand analysis framework is used that is motivated by the concept of summing up the loads of each appliance separately, in which case levels are the appropriate regressand and temperature response is scaled by some measure of the size of a house.

First, I present a mathematical justification for the ln specification. Second, I present some results using levels as the regressand after making appropriate adjustments. The results across vintage are similar.

$$
\begin{aligned}
kwhperday_{it} &= base_i + heat_{it} + cool_{it} && (4.1)\\
kwhperday_{it} &= base_i + f(weather) \times f(size) \times f(other) && (4.2)\\
kwhperday_{it} &= base_i + Z && (4.3)\\
ln(kwhperday_{it}) &= ln(base_i) + \frac{1}{base_i} * Z && (4.4)\\
&\quad \text{via Taylor approximation around Z=0} &&\\
ln(kwhperday_{it}) &= ln(base_i) + \frac{1}{base_i} * f(weather) \times f(size) \times f(other) && (4.5)\\
&\quad \text{assuming } \frac{f(size_i)}{base_i} = Q\text{, a constant} &&\\
ln(kwhperday_{it}) &= ln(base_i) + Q * f(weather) \times f(other) && (4.6)
\end{aligned}
$$

The derivation above begins with a partition of energy use into a base usage that is temperature and time invariant followed by heating and cooling loads that vary by time through weather's variation over time. The next step takes the natural log and then expands via a Taylor expansion. Under the maintained hypothesis that a function of size enters multiplicatively and that the ratio of base usage to the function of size is constant, size can then be omitted. Intuitively, this specification assumes that percent changes of bills are the comparable metric across buildings of different size. The $f(other)$ term would include vintages, housing characteristics, and household characteristics. [1]

Alternatively, one could directly estimate Equation 4.2 by choosing a functional form for $f(weather) \times f(size) \times f(other)$ when such data is available at a fine spatial resolution. My data at the Zip9-level, which on average has 5-10 households, is spatially more disaggregated than most other data. Weather was parameterized as a function of CDD and HDD and its squares.

A natural assumption to make is that cooling and heating loads scale by size, so that $f(size) = sqft$. This turns out to not be a good assumption, as shown below. I first estimate

[1] A reasonable alternative approach would be to use Box-Cox transformations to estimate nonlinearly the impact of size and choosing the model with the best fit.

the cumulative temperature response across vintages without other controls as described in Equation 4.7.

$$\begin{aligned} kWh\_useperday_{ijt} &= \sum_{v=1}^{VINTAGES} V_{jv} * (\beta_{1v} SQFT \times CDD_{it} + \beta_{2v} SQFT \times CDD_{it}^2 + \\ &\quad +\beta_{3v} SQFT \times HDD_{it} + \beta_{4v} SQFT \times HDD_{it}^2) \\ &\quad +\alpha_i + \varepsilon_{it} \end{aligned} \tag{4.7}$$

The results in Figure 1.20 show that new buildings are much less temperature responsive, contrary to other specifications. I then re-estimate this constrained to areas where the sqft variable is between 1300 and 1600 sqft which is a range of sqft with substantial overlap for all vintages. The results in Figure 1.21 show that new buildings perform worse, as is expected because they have much more air conditioning. The reason the two results differ is because the median of the sqft variable is larger for new houses and cooling and heating loads scale less than proportionately to sqft. Hence, the assumption that $f(size) = sqft$ overcorrects for size.[2]

While still using levels, I estimate a less functionally constrained version of $f(weather) \times f(size) \times f(other)$ in Equation 4.2. Size is restricted to sqft between 1300 and 1600. $f(size) = (\alpha_0 + \alpha_1 * sqft)$ which is a first order approximation applied to this narrow range of sqft. A similar first order approximation is used for air conditioning, and vintage is given by an indicator variable, similar to the main specification. The final specification has 64 parameter estimates.

$$\begin{aligned} kwhperday_{it} &= base_i + f(weather) \times f(size) \times f(other) \\ &\quad \text{where} \\ f(weather) &= \gamma_1 CDD + \gamma_2 CDD^2 + \gamma_3 HDD + \gamma_4 HDD^2 \\ f(size) &= \alpha_0 + \alpha_1 * sqft \\ f(other) &= \sum_{v=1}^{VINTAGES} \delta_v * VintageDummy_v * (\delta_0 + \delta_{CAC} * CentralAirConditioning) \end{aligned} \tag{4.8}$$

Figure 1.15 shows the results of the regression by predicting the value of electricity consumption $kwhperday_{it}$, for a reference 1500sqft house with central air conditioning for each vintage. Because of the large number of covariates, the regression results are omitted.

[2]KEMA-XENERGY (2004) models cooling load as scaling by external surface area. If a building doubles in size, the external surface area will less than double. For example, a cube on the ground has 5 external faces (one exposed to the ground), but two cubes side by side only have 8 external faces.

The results show that the 1990s and 1970s buildings may have lower temperature response after controlling for air conditioning and size, but that the difference is not statistically significant. Focusing just on the 1990s buildings, the range of the difference at 75°F is -2 to +1.5 kwhperday. This translates into an -8% to +6% difference in temperature response which is lower than the range given by the main specification.

4.1.3 Aggregation

The aggregation issue can be described by referring to the discussion of Blundell and Stoker (2005) which focuses on aggregation issues in demand systems and other scenarios. Aggregation presents biases when the underlying data generating process has cross-terms and there are non-zero covariances. For example, the following data generating process has no cross terms and could be estimated by data aggregated spatially across j.

$$
\begin{aligned}
y_{ij} &= \beta_0 + \beta_1 * x_{ij} + \beta_2 * z_{ij} + \varepsilon_{ij} && (4.9)\\
E_j[y_{ij}] &= \beta_0 + \beta_1 * E_j[x_{ij}] + \beta_2 * E_j[z_{ij}] + E_j[\varepsilon_{ij}] && (4.10)\\
y_j &= \beta_0 + \beta_1 * x_j + \beta_2 * z_j + \varepsilon_j && (4.11)
\end{aligned}
$$

In the presence of a cross term, the aggregation presents bias if there are covariances. In the example below, the relationship between the individual level coefficient, β_3, and the aggregate regression parameter, γ_3, is $\beta_3 = \gamma_3 \times \frac{E_j[x_{ij}]*E[z_{ij}]}{E_j[x_{ij}*z_{ij}]}$. The two equal if an only if the covariance, $Cov(x_{ij}, y_{ij})$, is zero.

$$
\begin{aligned}
y_{ij} &= \beta_0 + \beta_3 * x_{ij} * z_{ij} + \varepsilon_{ij} && (4.12)\\
E_j[y_{ij}] &= \beta_0 + \beta_3 * E_j[x_{ij} * z_{ij}] + E_j[\varepsilon_{ij}] && (4.13)\\
E_j[y_{ij}] &= \beta_0 + \beta_3 * (E_j[x_{ij}] * E[z_{ij}] + Cov(x_{ij}, y_{ij})) + E_j[\varepsilon_{ij}] && (4.14)\\
y_j &= \beta_0 + \gamma_3 * x_j * z_j + \varepsilon_j && (4.15)
\end{aligned}
$$

Aggregation problems are less likely with county assessor's data than with census block group data. County assessor's data is matched at the Zip9-level, which is about 5-10 households. Hence, it is hoped that covariates in a Zip9-level are relatively homogeneous in terms of house size, vintage of year built, and ownership of air conditioning. Census block groups, at 300-700 households each are much more likely to have these issues.

I have not done aggregation of bill to the census block or zip code level. Aggregation of all bills within a census block group can be done only if the panel is balanced; otherwise some bills exist in some years but not in others. A large proportion of properties have occupant turnover. If occupant turnover were random, dropping unbalanced observations would not present bias, but it is plausible that certain homes are more likely to have occupant turnover.

4.1.4 Extended Data Discussion

There are two datasets depending on the building characteristic information used. The first dataset uses ZIP9-level data from county assessor's information. The second dataset uses census block group-level data from the 2000 Census.

The billing data was cleaned. Bills with 25 days or less or 35 days or more were dropped (about 5%). Bills with less than 2kWh/day or more than 80kWh/day are outliers were also dropped (about 4%).

For the ZIP9 data, assessor's data primarily includes complete records of square footage, year built, and air conditioning ownership for single family homes. Records were dropped if there was more than 10 bedrooms, square footage less than 200 or greater than 10000, missing ZIP code, or the structure was built before 1850 or after 2000. Many of these were obvious data errors because they contained internally inconsistent values, such as many bedrooms but very little square footage. Census block group information was used to identify areas where more than 95% of the households were in single family structures and decreases the sample those areas that satisfy these criteria. Next, at the ZIP9-level, the proportion of houses with central air conditioning, the median structure size, and the proportion of buildings built in each vintage category were attributed to each bill in that associated ZIP9.

For the census block group data, a 1-in-5 subsample of observations was used to enable the estimation to be run on a Linux server with 8GB of RAM and an Intel Quadcore processor, running Stata 10.0 MP.

The spatial matching of weather, census block groups, and ZIP9s merits some description. Weather data is available on a 4km x 4km grid. Census block groups are given as polygons. ZIP9s are given as points, but the ZIP9 are ranges of street addresses. Typically opposite sides of the street will have different ZIP9s. To describe the matching from the perspective of the bill, the bill's ZIP9 is matched to the census block group and 4km by 4km grid square that contains the Zip9 point.

Chapter 5

References

Abrishami, Mohsen, Sylvia Bender, Kae C. Lewis, Nahid Movassagh, Peter Puglia, Glen Sharp, Kate Sullivan, Mitch Tian, Belen Valencia, and David Videvar, "Energy Demand Forecast Method Report: Companion Report to the California Energy Demand 2006-2016 Staff Energy Demand Forecast Report," Technical Report June 2005.

Acemoglu, D. and S. Johnson, "Disease and Development: The Effect of Life Expectancy on Economic Growth," *Journal of political economy*, 2007, *115* (6), 925–985.

Alesina, A. and B. Weder, "Do Corrupt Governments Receive Less Foreign Aid?," *The American Economic Review*, 2002, *92* (4), 1126–1137.

____ **and D. Dollar**, "Who gives foreign aid to whom and why?," *Journal of economic growth*, 2000, *5* (1), 33–63.

Anand, S. and A. Sen, "Human development Index: Methodology and Measurement," Human Development Report Office, United Nations Development Programme (UNDP), New York 1994.

____ **and** ____ , "Concepts of human development and poverty a multidimensional perspective," Human Development Papers, 1997 1997.

____ **and** ____ , "The Income Component of the Human Development Index," *Journal of Human Development*, 2000, *1* (1).

____ **and M. Ravallion**, "Human development in poor countries: on the role of private incomes and public services," *The Journal of Economic Perspectives*, 1993, *7* (1), 133–150.

Anderson, R.N., "A method for constructing complete annual US life tables," *Vital and health statistics. Series 2, Data evaluation and methods research*, 2000, (129), 1.

A new approach at Copenhagen, *Chinadialogue* 2892, Part1. Rutgers Climate and Social Policy Initiative 2009.

Arcelus, F.J., B. Sharma, and G. Srinivasan, "Foreign capital flows and the efficiency of the HDI dimensions," *Global Economy Journal*, 2005, *5* (2), 4.

Aroonruengsawat, Anin and Maximillian Auffhammer, "Impacts of Climate Change on Residential Electricity Consumption: Evidence From Billing Data," California Energy Commissoin, CEC-500-2009-018-D March 2009.

___ , ___ , and Alan Sanstad, "The impact of state level building codes on residential electricity consumption," Technical Report, UC Green Building Conference 2009.

Aturupane, H., P. Glewwe, and P. Isenman, "Poverty, human development, and growth: an emerging consensus?," *The American Economic Review*, 1994, *84* (2), 244–249.

Auffhammer, Maximillian, Carl Blumstein, and Meredith Fowlie, "Demand-Side Management and Energy Efficiency Revisited," *Energy Journal*, 2008, *29* (3).

Babylon, "Definition of the term Developing Country in Dictionary Babylon," http://www.babylon.com/definition/developing_countries/English (last accessed: 29 October 2009).

Baliamoune-Lutz, M., "On the Measurement of Human Well-being. Fuzzy Set Theory and Sens Capability Approach," 2004.

Bandyopadhyay, S., H.J. Wall, and Federal Reserve Bank of St. Louis, "The determinants of aid in the Post-Cold War era," *REVIEW-FEDERAL RESERVE BANK OF SAINT LOUIS*, 2007, *89* (6), 533.

Barro, RJ and J. Lee, "W (2001) International data on educational attainment: updates and implications," *Oxford Economic Papers*, *53* (541), 63.

Barro, R.J. and J.W. Lee, "International comparisons of educational attainment," *Journal of monetary Economics*, 1993, *32* (3), 363–394.

___ and X. Sala i Martin, "Convergence," *Journal of political Economy*, 1992, pp. 223–251.

Bate, R. and K. Boateng, "Drug Pricing and Its Discontents: At Home and Abroad," *Health Policy Outlook*, 2007.

Baxter, L.W. and K. Calandri, "Global warming and electricity demand:: A study of California," *Energy Policy*, 1992, *20* (3), 233–244.

Becker, Randy A and J Vernon Henderson, "Effect of Air Quality Regulations on Polluting Industries," *Jounral of Polical Economy*, 2000, *108*, 379–421.

Betz, R., W. Eichhammer, and J. Schleich, "Designing national allocation plans for EU-emissions trading- a first analysis of the outcomes," *Energy & Environment*, 2004, *15* (3), 375–426.

Blanchflower, D.G. and A.J. Oswald, "Happiness and the Human Development Index: the paradox of Australia," *Australian Economic Review*, 2005, *38* (No. 3), 307–318.

Blundell, Richard and Thomas Stoker, "Heterogeneity and Aggregation," *Journal of Economic Literature*, 2005, *43* (2), 347–391.

Bovenberg, A. and L.H. Goulder, *Behavioral and Distributional Effects of Environmental Policy*, Chicago: University of Chicago Press,

Brown, S.J. and J.B. Warner, "Using daily stock returns: The case of event studies," *Journal of Financial Economics*, 1985, *14* (1), 3–31.

Building Codes Assistance Project, "Code Status," bcap-energy.org. accessed 7/30/2009, archived.

Burtraw, D. and K. Palmer, "Compensation rules for climate policy in the electricity sector," *Journal of Policy Analysis and Management*, 2008, *27* (4), 819–847.

California Energy Commission, "Integrated Energy Policy Report," Technical Report, California Energy Commission 2007.

Charityscorecard.org, http.//www.charityscorecard.org/ (last accessed 2 January, 2009) .

Convery, F., D. Ellerman, and C. De Perthuis, "The European Carbon Market in Action: Lessons from the First Trading Period," Report 158. MIT Joint Program on the Science and Policy of Global Change 2008.

Costa, Dora L. and Matthew E. Kahn, "Why Has California's Residential Electricity Consumption Been So Flat since the 1980s?: A Microeconometric Approach," working paper #15978, NBER May 2010.

County of Riverside Assessor's Office, "Property Characteristics - Full; Catalog Code ASPRCHAR," CD-ROM Aug 2010.

Daily Times, "Pakistan ranked at 135th in human development," *Daily Times*, September 15 2005.

Dasgupta, P., "Valuing objects and evaluating policies in imperfect economies," *The Economic Journal*, 2001, *111* (471), 1–29.

Deaton, A.S. and A. Heston, "Understanding PPPs and PPP-based national accounts," National Bureau of Economic Research Cambridge, Mass., Working Paper No 14499 2008.

Definition of the term Developing Country in Dictionary "Wikipedia", http://en.wikipedia.org/w/index.php?title= Developing_country&oldid=188947979 (last accessed 1 February 2008).

Delarue, E., K. Voorspools, and W. Dhaeseleer, "Fuel switching in the electricity sector under the EU ETS: review and prospective," *Journal of Energy Engineering*, 2008, *134*, 40.

Desai, M., "Human development:: Concepts and measurement," *European Economic Review*, 1991, *35* (2-3), 350–357.

Deschênes, Olivier and Michael Greenstone, "Climate Change, Mortality and Adaptation: Evidence from Annual Fluctations in Weather in the U.S.," Massachusetts Institute of Technology Department of Economics Working Paper Series Dec 2008. Working Paper 07-19.

Dowrick, S. and J. Quiggin, "True measures of GDP and convergence," *The American Economic Review*, 1997, *87* (1), 41–64.

Easterlin, R.A., "The worldwide standard of living since 1800," *The Journal of Economic Perspectives*, 2000, *14* (1), 7–26.

Easterly, W., R. Levine, and D. Roodman, "Aid, policies, and growth: comment," *The American Economic Review*, 2004, *94* (3), 774–780.

Ellerman, A.D. and B.K. Buchner, "The European Union emissions trading scheme: origins, allocation, and early results," *Review of Environmental Economics and Policy*, 2007, *1* (1), 66.

____ , P.L. Joskow, and Pew Center on Global Climate Change, "The European Union's Emissions Trading System in Perspective," 2008.

Energy Information Administration, "State Energy Data System," www.eia.doe.gov/emeu/states/_seds.html 2007. accessed 7/30/2009, local archive from 2007.

____ , "Residential Energy Consumption Survey (RECS), 2005," www.eia.doe.gov/emeu/recs/ 2009.

Environmental Protection Agency, "2010 U.S. Greenhouse Gas Inventory Report," Technical Report, U.S. Environmental Protection Agency 2010.

Fama, E.F., L. Fisher, M.C. Jensen, and R. Roll, "The adjustment of stock prices to new information," *International Economic Review*, 1969, *10* (1), 1–21.

Global, Geneva, "Geneva Global Performance Philanthropy," http.//www.genevaglobal.com/sector-priorities/high-hd/ (last accessed 2 January 2009), Wayna,PA 2007.

Globerman, Steven and Daniel Shapiro, "Global Foreign Direct Investment Flows: The Role of Governance Infrastructure," *World Development*, 2002, *30* (11), 1899 – 1919.

Goulder, L.H., M.A.C. Hafstead, and M. Dworsky, "Impacts of alternative emissions allowance allocation methods under a Federal Cap-And-Trade Program," *Journal of Environmental Economics and Management*, 2010.

Gray, W.B., "The cost of regulation: OSHA, EPA and the productivity slowdown," *The American Economic Review*, 1987, *77* (5), 998–1006.

____ **and R.J. Shadbegian**, "Environmental regulation, investment timing, and technology choice," *The Journal of Industrial Economics*, 1998, *46* (2), 235–256.

Greening, Lorna A., David L. Greene, and Carmen Difiglio, "Energy efficiency and consumption – the rebound effect – a survey," *Energy Policy*, 2000, *28* (6-7), 389 – 401.

Guindon, G.E. and D Boisclair, "Past, current and future trends in tobacco use," WHO Tobacco Control Papers, TRENDS2003 2003.

Hargittai, E., "Holes in the Net. The Internet and International Stratification Revisited," 1998.

Hirst, Eric, "Progress and Potential in Evaluating Energy Efficiency Programs," *Evaluation Review*, 1990, *14* (1), 192–205.

Horowitz, Marvin J., "Changes in electricity demand in the United States from the 1970s to 2003," *Energy Journal*, 2007, *28* (3), 93–119.

Human Development Report, United Nations Development Program, Oxford University Press 1990-2006.

i Martin, X. Sala, "The World Distribution of Income: Falling Poverty and Convergence, Period*," *Quarterly Journal of Economics*, 2006, *121* (2), 351–397.

Jacobsen, Grant D and Matthew J Kotchen, "Are Building Codes Effective at Saving Energy? Evidence From Residential Billing Data in Florida," Technical Report 2009.

Joskow, Paul L. and Donald B. Marron, "What Does a Negawatt Really Cost? Evidence from Utility Conservation Programs," *Energy Journal*, 1992, *13* (4), 41–74.

___ **and** ___ , "What Does Utility-Subsidized Energy Efficiency Really Cost?," *Science*, 1993, *260* (5106), 281–370.

Kahn, S. and C.R. Knittel, "The Impact of the Clean Air Act Amendments of 1990 on Electric Utilities and Coal Mines: Evidence from the Stock Market," working paper CSEMWP-118, University of California Energy Institute 2003.

Keiser, J., J. Utzinger, M. Tanner, and B.H. Singer, "Representation of authors and editors from countries with different human development indexes in the leading literature on tropical medicine: survey of current evidence," *British Medical Journal*, 2004, *328* (7450), 1229.

Kelley, A.C., "The Human Development Index:" Handle with Care"," *Population and Development Review*, 1991, *17* (2), 315–324.

KEMA-XENERGY, "California Statewide Residential Appliance Saturation Survey: Final Report," Technical Report, California Energy Commission June 2004.

Kettner, C., A. Koppl, S.P. Schleicher, and G. Thenius, "Stringency and distribution in the EU Emissions Trading Scheme: first evidence," *Climate Policy*, 2008, *8* (1), 41–61.

Krueger, A.B. and M. Lindahl, "Education for Growth: Why and For Whom?," *Journal of Economic Literature*, 2001, *39*, 1101–1136.

la Fuente, A. De and R. Doménech, "Human capital in growth regressions: how much difference does data quality make?," *Journal of the European Economic Association*, 2006, *4* (1), 1–36.

Larsen, Bodil Merethe and Runa Nesbakken, "Household electricity end-use consumption: results from econometric and engineering models," *Energy Economics*, 2004, *26* (2), 179–200.

Leigh, A. and J. Wolfers, "Happiness and the Human Development Index: Australia Is Not a Paradox," *Australian Economic Review*, 2006, *39* (2), 176–184.

Linn, J., "Stock prices and the cost of environmental regulation," 06-011 WP, MIT Center for Energy and Environmental Policy Research 2006.

Linn, Joshua, "The effect of cap-and-trade programs on firms' profits: Evidence from the Nitrogen Oxides Budget Trading Program," *Journal of Environmental Economics and Management*, 2010, *59* (1), 1 – 14.

List, J.A., D.L. Millimet, and W. McHone, "The unintended disincentive in the Clean Air Act," *The BE Journal of Economic Analysis & Policy*, 2004, *4* (2), 2.

Loughran, David S. and Jonathan Kulick, "Demand-side management and energy efficiency in the United States," *The Energy Journal*, 2004, *25* (1), 19–43.

MacKinlay, A.C., "Event studies in economics and finance," *Journal of economic literature*, 1997, *35* (1), 13–39.

Marshall, Lynn and Tom Gorin, "California energy demand 2008-2018 staff revised forecast," Technical Report November 2007.

Mazumdar, K., "A note on cross-country divergence in standard of living," *Applied Economics Letters*, 2002, *9* (2), 87–90.

McGillivray, M., "The Human Development Index: yet another redundant composite development indicator?," *World Development*, 1991, *19* (10), 1461–1468.

Measurement of Human Development. Seven Questions, Lecture Reading by Deputy Director of the Human Development Report Office, UNDP 2000.

Moreno-Ternero, J.D. and J.E. Roemer, "Impartiality, priority, and solidarity in the theory of justice," *Econometrica*, 2006, *74* (5), 1419–1427.

Morgenstern, O, *On the Accuracy of Economic Observations*, second edition ed., Princeton University Press, 1970.

Morse, S., "Greening the United Nations' Human Development Index?," *Sustainable Development*, 2003, *11* (4), 183–198.

Nadel, Steven M. and Kenneth M. Keating, "Engineering Estimates versus Impact Evaluation Results: How Do They Compare and Why?," in "Energy Program Evaluation: Uses, Methods, and Results, Proceedings of the 1991 International Energy Program Evaluation Conference" 1991.

Nations, United, UN-NADAF Midterm Review, September 1996 1996.

Neary, J.P., "Rationalizing the Penn World Table: true multilateral indices for international comparisons of real income," *The American Economic Review*, 2004, *94* (5), 1411–1428.

Neumayer, E., "The determinants of aid allocation by regional multilateral development banks and United Nations agencies," *International Studies Quarterly*, 2003, *47* (1), 101–122.

Noorbakhsh, F., "The human development index: some technical issues and alternative indices," *Journal of International Development*, 1998, *10* (5), 589–605.

——, "International convergence or higher inequality in Human Development, evidence for 1975 to 2002," *WIDER Research Paper*, 2006, *15.*

Oberndorfer, U., "EU Emission Allowances and the stock market: Evidence from the electricity industry," *Ecological Economics*, 2009, *68* (4), 1116–1126.

Ogwang, T., "Inter–country inequality in human development indicators," *Applied Economics Letters*, 2000, *7* (7), 443–446.

O'Neill, H., "Ireland's Foreign Aid in 2004," *Irish Studies in International Affairs*, 2005, *16* (-1), 279–316.

Oswald, A.J., "Happiness and economic performance," *The Economic Journal*, 1997, *107* (445), 1815–1831.

Pakistan Times, "Pakistan improves Human Development Indicators," *Pakistan Times National News Desk*, July 20 2004.

Peoples Daily, "Ghana's Human Development Performance Recommended by UNDP," *Peoples Daily*, July 11 2001.

Petersen, M. and L. Rother, "Maker Agrees to Cut Price of 2 AIDS Drugs in Brazil," *New York Times*, March 31 2001.

Pillarisetti, J.R., "An empirical note on inequality in the world development indicators," *Applied Economics Letters*, 1997, *4* (3), 145–147.

Puller, S.L., "The strategic use of innovation to influence regulatory standards," *Journal of Environmental Economics and Management*, 2006, *52* (3), 690–706.

Quah, D.T., "Twin peaks: growth and convergence in models of distribution dynamics," *The Economic Journal*, 1996, *106* (437), 1045–1055.

RAND California, "County population estimates," http://ca.rand.org/stats/statlist.html 2010. accessed Sept 23, 2010.

Reiss, Peter C and Matthew W. White, "What changes energy consumption? Prices and public pressures," *RAND Journal of Economics*, 2008, *39* (3ei), 636–663.

Rogoff, K., "The purchasing power parity puzzle," *Journal of Economic literature*, 1996, *34* (2), 647–668.

Ryan, S., “The costs of environmental regulation in a concentrated industry,” working paper, MIT Center for Energy and Environmental Policy Research 2005.

Salop, S.C. and D.T. Scheffman, “Raising rivals’ costs,” *The American Economic Review*, 1983, *73* (2), 267–271.

Sanyal, R.N. and S.K. Samanta, “Determinants of bribery in international business,” *Thunderbird International Business Review*, 2004, *46* (2), 133–148.

Schlenker, Wolfram and Michael Roberts, “Nonlinear Temperature Effects indicate Severe Damages to U.S. Crop Yields under Climate Change,” *Proceedings of the National Academy of Sciences*, 2009, *106* (37), 15594–15598.

Seade, J., “Profitable cost increases and the shifting of taxation: equilibrium response of markets in Oligopoly,” *The Warwick Economics Research Paper Series (TWERPS)*, 1985.

SearchWiki, “Definition of the term Developing Country in Dictionary SearchWiki,” http://www.searchthewiki.com/wikisearch/?q=Developing%20-countries (last accessed: 29 October 2009) 2009.

Sen, A., “On weights and measures: informational constraints in social welfare analysis,” *Econometrica*, 1977, pp. 1539–1572.

——, “The living standard,” *Oxford Economic Papers*, 1984, *36*, 74–90.

——, *Commodities and Capabilities*, North-Holland, Amsterdam, 1985.

——, “A decade of human development,” *Journal of Human Development*, 2000, *1* (1), 17–23.

Sen, A.K., “On Ethics and Economics, The Royer Lectures,” 1987.

Sijm, J., K. Neuhoff, and Y. Chen, “CO2 cost pass-through and windfall profits in the power sector,” *Climate Policy*, 2006, *6* (1), 49–72.

Smale, R., M. Hartley, C. Hepburn, J. Ward, and M. Grubb, “The impact of CO2 emissions trading on firm profits and market prices,” *Climate Policy*, 2006, *6* (1), 31–48.

Srinivasan, T.N., “Human development: a new paradigm or reinvention of the wheel?,” *The American Economic Review*, 1994, *84* (2), 238–243.

Sudarshan, A. and J. Sweeney, “Deconstructing the “Rosenfeld Curve”,” Stanford, CA: Stanford University. Online at piee.stanford.edu/cgibin/docs/publications/sweeney/Deconstructing the Rosenfeld Curve. pdf 2008.

United States Census Bureau, "Census 2000 Summary File 3 ASCII text data files," http://www.census.gov/support/SF3ASCII.html July 2009. accessed July 2009.

Veith, S., J.R. Werner, and J. Zimmermann, "Capital market response to emission rights returns: Evidence from the European power sector," *Energy economics*, 2009, *31* (4), 605–613.

Weiner, C., "The impact of industry classification schemes on financial research," *SFB 649 Discussion Papers*, 2005.