Forecasting US CO_2 Emissions Using State-Level Data

Ralf Steinhauser and Maximilian Auffhammer*

University of California Berkeley

Draft from May 2005

Abstract

This paper tests the out of sample predictive ability of reduced form models found in the literature on forecasting CO_2 emissions. We show that for a newly available panel data set covering the fifty U.S. States and Washington D.C. during the years 1960 - 2000, the benchmark models found in the literature are outperformed by over 3000 of the considered models. A search over a large universe of models selects a "best" performing model using the mean square forecast error from a simulated out of sample forecast experiment. Forecasts of aggregate CO_2 emissions for the United States are provided.

Keywords: Forecasting, Climate Change, United States, CO_2 Emissions, Data Snooping

*Send correspondences to Department of Agricultural and Resource Economics, 310 Giannini Hall, University of California, Berkeley, CA 94720-3310, USA, Phone: (510) 642-3345, Fax: (510) 643-8911, E-Mail: steinhauser@are.berkeley.edu or auffhammer@are.berkeley.edu.

1. Introduction

The possibility of global climate change and its consequences pose a significant environmental threat for mankind over the next century and beyond. The Intergovernmental Panel on Climate Change (IPCC, 2001) predicts an increase in globally averaged surface temperatures between 1.4 to 5.8 °C by the end of the current century[1]. The largest degree of uncertainty is over the regional impacts of this global warming trend, not over the trend in mean global temperatures. The main trace gas from anthropogenic sources partially responsible for this warming is CO_2, which accounts for about 50% of the radiative forcing and its share is predicted to rise to 2/3 over this century (IPCC, 2001).[2] Its average atmospheric life span of 125 years (Frey, Staehelin-Witt and Bloechlinger, 1991) leads us to follow the literature in considering it as a stock pollutant.

Forecasts of future trace gas emissions levels serve two main purposes. First, they are used as an important input for global climate models (GCM), which are used to predict warming and precipitation trends. Second, individual countries use business as usual (BAU) forecasts to calculate expected costs of emission reductions in future periods. Since anthropogenic CO_2 emissions in industrialized countries come from the combustion of fossil fuels, potential reductions in emissions are closely tied to economic activity. Producing optimal forecasts of global and country level emissions therefore has a first order effect on predicted warming trends as well as the optimal policy decisions made (*i.e.* whether to ratify the Kyoto agreement). Further, incorrect predictions of trends in warming and precipitation due to suboptimal forecasts of emissions will result in biased calculations of the benefits from preventing global warming and therefore have a second order effect on the policy adoption decisions.

BAU emissions forecasts of CO_2 from anthropogenic sources on a country level are calculated using a variety of models. The economics literature can be divided into two general strands. Schmalensee, Stoker and Judson (1998) and Holtz-Eakin and Selden (1995) use reduced form models to calculate forecasts of aggregate global CO_2 emissions using a panel of country level emissions. These models focus on the functional form of the dependence of per capita emissions and income. The empirical finding of negative marginal propensity to emit at high levels of income leads Schmalensee et al. (1998) to provide forecasts consistent with

[1] A recent large scale study by Stainforth, Aina, Christensen, Collins, Faull, Frame, Kettleborough, S. Knight, Murphy, Piani, Sexton, Smith, Spicer, Thorpe and Allen (2005) suggests likely increases between 2 and 11 °C.

[2] The main greenhouse gases (GHGs) in addition to CO_2 are methane (CH_4), nitrogen oxides (NO_x), nitrous oxide (N_2O), ozone (O_3) and halocarbons like $CFCL_3$ and CF_2CL_2.

an Environmental Kuznets Curve relationship, whereby per capita emissions increase at low levels of income and decrease after a threshold level of income is reached. Holtz-Eakin and Selden (1995) using a very similar specification do not find empirical evidence supporting an EKC and provide aggregate forecasts which are significantly lower. The second strand of the economics literature adopts a more structural approach by using aggregate computable general equilibrium (CGE) models (Garbaccio, Ho and Jorgenson, 1999). This approach models emissions from individual sectors of the economy and explicitly takes into account effects of population growth, capital accumulation, technological change, and changing patterns of demand. Aside from producing forecasts these models are well suited to study the impact of policy measures such as carbon taxes.

The engineering literature can also be divided into two similar strands. The first consists of reduced form statistical decomposition models (Yang and Schneider, 1998), which are similar to the reduced form economics literature but pay more attention to modeling technological change instead of functional form of the emissions income relationship. The second strand of the engineering literature employs a number of large scale general equilibrium simulation models such as the Energy Information Agency's National Energy Modeling System (NEMS)[3]. These models are characterized by a tremendous amount of detail at the sector level as well as the modeling of technology.

None of the models above use an explicit out-of-sample model selection process, but are parameterized according to in sample fit.[4] Further, for the reduced form approach, the models appear to be selected by conducting a model selection search over a small space of models. These two issues are addressed in this paper by introducing an extensive systematic model selection approach to the CO_2 forecasting literature.

The paper is divided into two main parts. In the first part we will test the performance of the models proposed in the literature against the performance of a large universe of reduced form models. Estimation results suggest that the simplistic reduced form models found in the literature are outperformed by a large space of the considered model universe.

The second part of the paper provides an attempt to select a forecasting model from the model universe using mean square forecast error (MSFE) as a model selection criterion. For the entire paper we adopt an explicit forecasting approach to model selection, by making out-of-sample predictions based on information available in sample and selecting the best

[3]For an in depth description of the NEMS model, refer to http://www.eia.doe.gov/oiaf/aeo/overview/

[4]The EIA conducts an annual evaluation of its forecasts, but it is not clear from the literature how this impacts the reparametrization of the NEMS model. For a discussion of the NEMS model forecast evaluation see ONeill and Desai (2005).

performing model, similar to the simulated out-of-sample forecast evaluation in Watson, Marcellino and Stock (2003).

We conduct our tests of forecasting performance as well as the model selection search using a newly available state-level panel data set for U.S. CO_2 emissions covering the years 1960 - 2000. The focus of this paper is to select the reduced form model with the best out-of-sample forecasting performance from our model universe, rather than attempting to explain the mechanisms driving historical in sample variation, which we are currently attempting in a parallel research project. We use the United States as the country of study, since jointly they are the largest emitter of CO_2 and have not yet ratified the Kyoto accord. Providing accurate forecasts of US emissions is therefore important from a GCM as well as a policy perspective.

The next section provides a brief review of the literature. Section 3 discusses the test of benchmark models against our universe of models. Section 4 provides our model selection approach and forecasts from our "best" model. Section 5 concludes.

2. Background and Literature Review

There are two distinctly different approaches to modelling emissions of CO_2. The first is a structural modeling approach in which a set of free parameters in large scale general and partial equilibrium models are fixed by judgment and calibration. The second approach is using reduced form models with far fewer parameters which are solely calibrated based on the goodness of fit to historical data.

Structural modeling is the predominant approach in the natural science and engineering literature. The organizing framework of these models is based on the $I = P \cdot A \cdot T$ identity (Ehrlich and Holdren, 1971), where I stands for impact (emissions), P stands for population, A for the affluence (per capita GDP or income), and T for a technology index. The simplistic IPAT models imply that emissions monotonically increase with population and affluence and decrease with technological progress. The engineering literature has concentrated on modeling the technological change component of the IPAT model. This has involved the fine tuning of structural parameters to accurately emulate real data. A simple statistical decomposition approach is provided by Yang and Schneider (1998). More involved large scale simulation models underly the quasi-official IPCC reports as well as forecasts provided by the US Energy Information Administration (EIA)[5].

[5]The EIA uses the NEMS model to forecast the national energy system and CO_2 emissions, which are published in their Annual Energy Outlook. Since these forecasts are provided by an agency of the United

The economics literature has focused on reduced form models. The forecasting literature is based on early work by Grossman and Krueger (1993) and Selden and Song (1994) who look at the relationship between air pollutants and income. Shafik (1994), Holtz-Eakin and Selden (1995) and Grossman and Krueger (1995) are further examples of this reduced form approach. The focus of this literature is the empirical finding of an inverse-U relationship between emissions and/or ambient concentrations of pollutants and per capita income, which is known as the Environmental Kuznets Curve (EKC) hypothesis. The existence of such a relationship has been a the center of the debate surrounding future trends in emissions of local and global pollutants from developing and developed countries alike.

As Arrow et al. (1995) point out, the empirical finding of a Kuznets curve relationship does not provide any insight into what the driving factors behind the possible downturn are. Further, the findings for an EKC relationship are more robust for local air and water pollutants than for the global stock pollutant CO_2. Barbier (1997) provides an overview of the early EKC literature. Lieb (2005) provides a more recent overview of the specifications and empirical findings. Recent work by Millimet, List and Stengos (2003) and Levinson, Harbaugh and Wilson (2002) cast doubt on the robustness of the EKC specification. Schmalensee et al. (1998) use this framework to forecast emissions of CO_2 out of sample. They use a flexible version of the Environmental Kuznets Curve specification to forecast emissions. Holtz-Eakin and Selden (1995) before them used a simple quadratic income term to implement such an inverse-U relationship for CO_2 emissions. Both paper use the same source of data, although the latter observe emissions over a shorter time period. The provided forecasts differ substantially. For the year 2020 Schmalensee et al. (1998) predict emissions of 2.12 billion metric tons of carbon (average over the 6 scenarios), whereas Holtz-Eakin and Selden (1995) predict emissions of 1.26 billion tons of carbon. Both papers use in-sample model criteria to select their forecasting model, which may lead to suboptimal out of sample performance (McCracken and West, 2004).

In the literature on forecasting CO_2 emissions is no agreement on the existence of a negative marginal propensity to emit at high incomes, which is the central question of the debate on whether a turning point exists for gases with no direct local health impacts. Holtz-Eakin and Selden (1995) only found a diminishing marginal propensity to emit at the highest levels of income, while Schmalensee et al. (1998) find a turning point in the pollution income relationship.

The major advantage of the reduced form models from a practical perspective is that

States government, they are considered the official forecasts. The IPCC, for political reasons, does not publish country level forecasts.

they have lower data requirements, which allow for longer time series and facilitate analysis for countries where the structural approach is infeasible. They further do not require a large number of parametric assumptions be made, as do their structural counterparts. When using a reduced form approach to forecast, the assumptions about the functional form and the determination of the to be included variables becomes the central most important step. Therefor models are selected by an extensive specification search. The literature on forecasting model selection is vast, but can be divided into three approaches. The first and most commonly practiced approach is a sequential model selection approach, by which one starts with a general unrestricted model (GUM) based on the largest set of potential regressors and then selects the "best" forecasting model by sequential testing of zero parameter restrictions. In order to circumvent the issue of following a false path encompassing tests and model selection criteria are applied to select between competing models to obtain a a dominant minimal representation. Hoover and Perez (1999) provide an empirical implementation of this algorithm.[6] An alternate approach to constructing forecast models is the diffusion indexes approach of Stock and Watson (2002). For this method one constructs principal components from the space of covariates in a first stage, which are then included with lags of the dependent variable in a second stage. Model selection happens by using the SIC. Finally one could use Bayesian shrinkage estimation to a most general model and shrink the coefficients towards zero. In order to simulate a real forecasting scenario, we select our model by constructing 10 year ahead forecasts from a large universe of models and calculating the MSFE out of sample forecast following Watson et al. (2003). This approach in some sense, ignores potential structural breaks out-of-sample (i.e. the recession in the early 1990s). Pretending ignorance, we intend to discover which specification does best forecasting the level of per capita emissions.[7]

Since we only observe one realization of any time series, there is always the danger that the observed predictive power of the chosen model may be due to chance rather than true forecasting ability of the model. An additional problem is that most specification searches in practice are often not systematic nor can they hope to be comprehensive. This issue, commonly referred to as data snooping describes any situation in which data are used repeatedly for inference or model selection, but the reuse of the data is not accounted for in

[6]PC-GETS is a commercially available software package which implements an automated model selection procedure. (Giacomini and White, 2004) show that for a series of macroeconomic indicators, this approach does poorly compared to shrinkage and diffusion indices

[7]We are currently implementing each of these approaches to compare forecasts from the selected models by each approach to predictions from the optimal model using our out-of-sample experiment.

inference tests.[8,9] The reason for oversight of this issue in applied studies was the lack of an easily implementable and broadly applicable way of accounting for the impact of specification searches on inference tests.

White (2000) provided such a 'rigorously founded, generally applicable method for testing the null hypothesis that the best model encountered during a specification search has no predictive superiority over a benchmark model' and calls it the Reality Check. In this paper we attempt to account for the dual problems of ad hoc specification searches and data snooping. Following White (2000) we will systematically search over all models in a pre-defined universe and test the null hypothesis of no superiority of the best performing model over the benchmark model in a way which accounts for this search. This approach will significantly reduce the risk that we falsely claim to have found a model with good predictive power, within the bounds of our model universe.

3. Data and Model Selection

3.1 Data

Blasing, Broniak and Marland (2004) provide a new data set of CO_2 emissions for the 50 states and Washington D.C. for the years 1960 until 2000, which results in a balanced panel of 2091 observations. This is the longest and most complete CO_2 emissions data set for a single country at a sub-national level of aggregation. Blasing et al. (2004) used consumption data for coal, petroleum, and natural gas from the EIA State Energy Data Report to calculate carbon emissions.[10] The data do not account for carbon oxidized during gas flaring, from the calcining of limestone during manufacture of cement, or carbon from bunker fuels.[11,12] Emissions are reported in teragrams (million metric tons) of carbon.

Income data are taken from the Bureau of Economic Analysis. We inflate personal income

[8]More formally Aldous (1989, p.252) defines data snooping as the situation where you have a family of test statistics $T(a)$ whose null distribution is known for fixed a, but where you use the test statistic $T = T(a)$ for some a chosen using the data.

[9]The problem of data snooping has been long understood and was, for example, pointed out by Cowles (1933) and Leamer (1978). More recently it has been brought to wide attention by Lo and MacKinley (1990).

[10]For petroleum, the data include energy production and transportation in each state, as well as oxidized carbon emissions from other end uses such as the production of plastics, fabrics or lubricants.

[11]Those neglected sources together account for only approximately 4% of the total carbon emissions, see Holtz-Eakin and Selden (1995).

[12]Despite the omission of these latter sources, the emissions estimates tend to be slightly higher than their well known national counterparts, which are provided by Marland, Boden and Andres (2004). The difference in the estimates is due to the use of *actual* consumption data from the U.S. Department of Energy, instead of estimates based on 'production + imports - exports - changes in stocks'.

by state into real year 2000 US$. As a further control variable we use population density, calculated from the Blasing et al. (2004) population data and land area data from the US Department of Commerce. We have also collected a dummy variables for coastal states, oil or gas producing states and coal producing states.

As Watson et al. (2003) point out, forecasting aggregate series, by constructing forecasts at a more disaggregate level of observation may result in improved forecasts. In order to use the data as a cross panel there needs to be sufficient heterogeneity in the data across individual states, which provides sufficient cross sectional and time series variability to identify parameters.

Figure 1: Per Capita CO_2 Emissions and Annual Growth Rate, by State

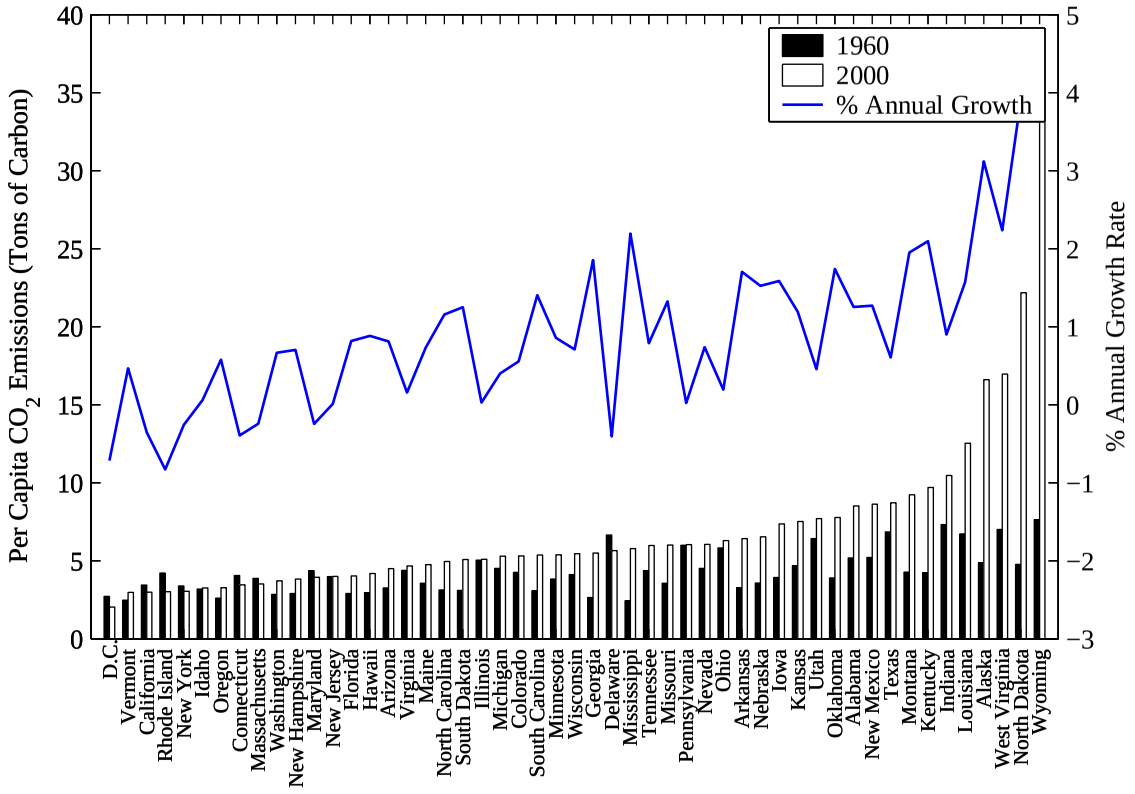


Figure 1 displays per capita emissions for 1960 and 2000 for all 50 states and Washington D.C. and shows that there is large heterogeneity in emissions data across states and time. The same is true for the income data, there is sufficient variation across states, with an average per capita income over all the 41 years and 51 regions of $21,730 and a standard deviation of $5,230. Annual income varies from $8,115 for Mississippi in 1960 to $42,116 for the District of Columbia in the year 1999, which is a 5 fold difference. There are also large

variations within a particular year. For example in 2000 the poorest state, Mississippi, had an average income of $21,034 which is only 50% of that of the richest state, Connecticut, with $41,570. An additional advantage compared to cross country panel Data is, that all series employed were measured with consistent definitions and units of observation, which decreases the likelihood of these factors introducing bias into parameter estimates.

3.2 The Reality Check Bootstrap

Traditional out of sample prediction tests such as Diebold and Mariano (1995) and West (1996) allow for the comparison of two competing models under different sets of assumptions. However, these and other traditional predictive ability tests ignore dependence between the results for different forecasting models. White (2000) provides a test which incorporates the dependence of results across forecasting models into a comprehensive bootstrap based test. Since this methodology has not been applied in the Environmental Economics literature, we provide a somewhat detailed description of the test below. The Reality Check (RC) tests the null hypothesis that the best model encountered in the specification search has no predictive superiority over a given benchmark model, taking the initial search over models into account. We therefore account for the possibility that the best performing model is selected by chance. To obtain the distribution of the best model's performance criterion, we use a bootstrap approach incorporating all models and get a p-value which provides an objective measure of how consistent the results of the best model are in relation to the sampling variation of the searched universe.

First, in order to compare the predictive ability among models, one needs to choose a selection criterion. Compared to the common in-sample measures like R^2, joint significants test or the Akaike Information Criteria, our goal is to choose a model with superior out-of-sample predictive ability. Therefore we use the MSFE of 10 periods out-of-sample forecasts as our selection criteria. To make the result less sensitive to certain forecasting periods we do 8 forecast starting 1983,1984,...,1998,1990. Each of the rolling windows forecasts $\tau = 10$ periods ahead using the information from the data up to that given period.

To introduce notation, a general reduced form that could determine an existing relationship between per capita carbon emissions c and some predictors X is $c_{i,t} = X_{i,t}\beta_{i,t} + \epsilon_{i,t}$, where β represents a vector of covariates and ϵ random error term. The subscript t stands for the time period and i for the state. If we now want to forecast emissions from 1989 to 1999, we use all data from year 1960 to 1989 to estimate $\hat{\beta}_{i,k,1989}$ with least squares and to predict

$\hat{X}_{i,k,1999}$. With these we can calculate the predicted emissions, for the model specification k, as $\hat{X}_{i,k,t+\tau}\hat{\beta}_{i,k,t}$. For each model we get 8 emission forecasts for 1993 through 2000, which we compare to the actual realizations $c_{i,t+\tau}$. The selection criteria for each model k relative to a benchmark model is defined as:

$$\hat{f}_{k,t+\tau} = \sum_{i=1}^{51} -(c_{i,t+\tau} - \hat{X}_{i,k,t+\tau}\hat{\beta}_{i,k,t})^2 + (c_{i,t+\tau} - \hat{X}_{i,0,t+\tau}\hat{\beta}_{i,0,t})^2 \tag{1}$$

for $t = 1983, \ldots, 1990$ and for each $k = 1, \ldots, L$ of the $L = 21168$ models. $k = 0$ represents the benchmark model. The difference in sum of squared errors for a model and the benchmark for a given t are summed over all states in equation 1 to acquire a measure for relative divergence of emission predictions on US level. Since a smaller value for sum of squared errors indicates relatively better model performance, a negative $\hat{f}_{k,t+\tau}$ indicates that for period $t+\tau$ the k^{th} model performed worse than the benchmark. The null hypothesis of the RC takes the form:

$$H_0 : \max_{k=1,\ldots,L} E[f_k] \leq 0 \quad .$$

Under the null hypothesis, therefore, we expect the benchmark to outperform the best of all models contained in the universe. The alternative is that the best model is superior to the benchmark. We estimate the first moment of the relative performance measure as the MSFE over the n=8 prediction periods from equation 1, i.e. $E[f_k] = \bar{f}_k = n^{-1}\sum_t \hat{f}_{k,t+\tau}$. The best performing model (BPM) from the search can be obtained from:

$$k^* = argmax_{k=1,\ldots,L}\bar{f}_k \quad .$$

In order to test whether the BPM significantly outperform the benchmark, we need to know the distribution of the estimator of f_{k^*}, i.e. $\bar{f}_{k^*}$. As White (2000) points out, this is difficult to derive analytically since $\bar{f}_{k^*}$ is the result of a maximization over a function space. To overcome this problem, one can numerically approximate the distribution by Monte Carlo simulation or a bootstrap. Hansen (2004) points out that the Monte Carlo implementation is not possible when the number of models searched over is larger than the number of forecasts. This is clearly the case here, so we use the bootstrap implementation.

The bootstrap resamples the $\hat{f}_{k,t+\tau}$ to construct a distribution for $\bar{f}_{k^*}$ and obtain the

p-value of the test statistic. White (2000) proposes the following test statistic:

$$T^{RC} = \max_{k=1,\ldots,L} n^{1/2} \bar{f}_k \quad ,$$

We generate the bootstrap resamples using the stationary bootstrap of Politis and Romano (1994). The stationary bootstrap constructs resamples $\theta_{b,t}$, for $b = 1, \ldots, B$. We chose B=500 resamples each with $n = 8$ draws and since there is reason to believe that we have dependence between forecasting results we choose a fixed block length of 4.[13]

The resampled statistic is now computed as follows:

$$\bar{f}^*_{k,b} = n^{-1} \sum_t \hat{f}_{k,\theta_{b,t}+\tau} \qquad \forall\, b = 1, \ldots, B \quad .$$

These can be viewed as independent draws from the distribution of $\hat{f}_k$. White (2000) shows that the distribution of $\bar{f}_{k^*}$ is properly approximated by:

$$T^{RC*}_{u,b} = \max_{k=1,\ldots,L} n^{1/2} (\bar{f}^*_{k,b} - \bar{f}_k) \qquad \forall\, b = 1, \ldots, B \quad . \tag{2}$$

By subtracting $\bar{f}_k$ he centers the bootstrap variables about zero. He chooses this centering because it is associated with the most conservative interpretation of the null. It is based on the null least favorable for the alternative. Hansen (2004) argues that this centering is too conservative, and that it allows poorly performing models in the universe to have substantial influence on the distribution of the test statistic. The result, he claims, is an upward bias on the RC p-value p_u. The reason that poorly performing models are able to have a disproportionate influence on the test statistic is that, precisely because they perform worse on average than the benchmark, they have a negative value of $\bar{f}_k$. This means that while the well-performing models are centered by having their (positive) $\bar{f}_k$ subtracted, the poorly-performing models have the absolute value of their (negative) $\bar{f}_k$ added. If a poorly performing model has a high variance, it can happen that it's centered bootstrapped test statistic $T^{RC*}_{u,b}$ is much larger then T^{RC} and so drive up the p-value p_u.

Hansen (2004) illustrates this property using simulations. On this basis he suggests that a test statistic with superior properties can be achieved by not centering the poorly performing, high variance models around zero. The two alternative centering approaches suggested by

[13] A larger block length is not advisable for our bootstrap, since we have a limited number of observations to get the test statistic, a smaller block length on the other had makes outlying poor performing models dominate more (see discussion below). The results are fairly robust to the choice of block length. The choice of $n = 8$ is owed to the fact that we look for the distribution of the 8 period average $\bar{f}_k$.

Hansen are slightly modified given in equation 3 and 4.[14] In the first alternative the models that are excluded from the centering around zero are in essence those whose $\bar{f}_k$ is less than the negative of their bootstrap standard error:

$$T_{c,b}^{RC*} = \max_{k=1,\ldots,L} n^{1/2}(\bar{f}_{k,b}^{*} - \bar{f}_k \cdot \mathbf{1}_{\{\bar{f}_k \geq -A_k\}}) \qquad \forall\, b = 1, \ldots, B \tag{3}$$

where $A_k = 1/4n^{-1/4}\sqrt{\widehat{var}(n^{1/2}\bar{f}_k)}$ and $\widehat{var}(n^{1/2}\bar{f}_k) = B^{-1}\sum_b (n^{1/2}\bar{f}_{k,b}^{*} - n^{1/2}\bar{f}_k)^2$. Hansen's second suggestion reduces the influence for all models that are worse then the benchmark itself:

$$T_{l,b}^{RC*} = \max_{k=1,\ldots,L} n^{1/2}(\bar{f}_{k,b}^{*} - max(\bar{f}, 0)) \qquad \forall\, b = 1, \ldots, B \tag{4}$$

We will report results for both White's and the variation of Hansen's adjusted method.

Given the bootstrapped distribution for $\bar{f}_{k^*}$ from equation 2 we can calculate the RC p-value. We sort the values $T_{u,b}^{RC*}$ and denote them as an order statistic, $T_{u,1}^{RC*}, T_{u,2}^{RC*}, \ldots, T_{u,B}^{RC*}$. Now we find the N for which $T_{u,N}^{RC*} \leq T^{RC} < T_{u,N+1}^{RC*}$ and get the Bootstrap Reality Check p-value as:

$$p_u = 1 - N/B \quad .$$

The same procedure is used to get the alternative p-values p_c and p_l based on the distributions suggested by Hansen. As a fourth comparison we report the 'naive' p-value, where one does not take the specification search into account and just apply the bootstrap to the best model alone. The naive p-value serves as a first guideline, since it is by definition weakly smaller. In case the null hypothesis is already rejected by the naive p-value no further computation is needed.

3.3 The Model Universe

As discussed above, in White's RC we do the specification search over a predefined model universe. We have collected data on carbon emissions, income, population and population

[14] Hansen goes in his paper even further and weights the relative average squared prediction errors with their standard deviations. When implemented in that way the results are no longer directly comparable to White's RC, which is why we retreat to the modified version.

density and several dummy variables. The models considered in the paper take the general form:

$$\mathrm{c}_{it} = \beta_1 \mathrm{state}_i + \beta_2 \mathrm{time}_t + \beta_3 f(\mathrm{income}_{it}) + \beta_4 g(\mathrm{pdens}_{it}) + \beta_{5i} \mathrm{c}_{it-1} + \varepsilon_{it} \tag{5}$$

Where c_{it} are per capita carbon emissions for state i in time period t, state_i is a state fixed effect or qualitative dummies for state i, time_t is a time fixed effect or time trend for period t, $f(\cdot)$ and $g(\cdot)$ are a higher order polynomials or splines, income_{it} and pdens_{it} are per capita income and population density respectively for the state in that time period, c_{it-1} are state i's lagged carbon emissions and ε_{it} is the error term. Starting with this general form we let all possible characteristics of the model vary to get a total of 21168 different models. The functional form for income varies from just a linear income term up to a 5^{th} order polynomial, as well as a spline on income. The functional form of population density varies from a linear to a quadratic polynomial. We also include lagged regressors. Further variations are the inclusion or exclusion of state fixed effect or qualitative dummies, time fixed effects or a linear or logarithmic time trend and the lagged state specific carbon emissions term. All of these possible combinations are included in levels and logarithmic form to allow for a multiplicative data generating process of the regression equation.[15] By using this wide range of specifications including fixed effects and dummies we controlling for omitted variable bias given the limited data. The state dummies reflect differences across states which are time invariant such as industry mix, climate and resource endowment. The dummy variables on coal and oil or gas producing states control for the very energy intensive energy production process in those states for the models where they are included. The time dummies and time trend control for changes in technology, environmental policies or relevant taxes, global economic development or other shocks as well as change in taste that are not related with affluence. The time effects will partially be able to control for the unobserved oil prices changes, but limited given the high volatility of oil prices over the period in question and the huge influence on fuel consumption.[16] One may therefore argue, that the error term is not entirely exogenous, but we accept that weakness existing in the entire reduced form literature on emission forecasting and will not attempt to deal with it more in this paper than done above. Table 1 below is a summary description of the model universe.

[15]This model universe is by no means exausitve, even given the limited data, but it covers systematically an extensive range of specifications that are commonly considered for reduced form modeling in this area.

[16]State level data on oil prices is not available and even U.S. oil price data going back to 1960 could not be found by the authors.

Table 1: Model universe

	Base Models	Additional Regressors	Transformations
Variations	• income per capita (y) up to the 5th power • population density (p) up to the 2nd power • for example: – $y + p$ – $y + y^2 + y^3 + y^4 + y^5$ – $y + y^2 + y^3 + p + p^2$ • income spline – 3-10 knots	• state fixed effects • time fixed effects – with structural breaks (1971-1983) • linear time trend • ln(time trend) • state dummies – coastal – oil/gas producing – coal producing	• levels • logs • adding regressor lags (pdens_{it-1} and income_{it-1}) • adding lagged dependent variable (c_{it-1})
Combinations	25+17	2*11+3*7+7*11+6=126	8
Model Universe		$(25 + 17)/2 * 126 * 8 =$ **21168**	

3.4 The Benchmark Model

In White's Reality Check method one compares the best performing forecasting model to a benchmark. In an application, Sullivan, Timmermann and White (1999) compare a large number of technical trading rules to the benchmark of holding cash. The benchmark works as an orientation that makes the performance of the best model quantifiable. We choose two widely used specifications of the reduced form literature (Schmalensee et al., 1998; Holtz-Eakin and Selden, 1995)[17] and one very basic model form the structural literature (Yang and Schneider, 1998). Schmalensee et al. (1998) propose the following specification:

$$ln(c_{it}) = \beta_1 state_i + \beta_2 time_t + \beta_3 F(ln(income_{it})) + \epsilon_{it}$$

where the variables are the same as before and $F(\cdot)$ represents a piecewise linear function with 10 segments.[18]

The model specification used by Holtz-Eakin and Selden (1995) is:

$$ln(c_{it}) = \beta_1 state_i + \beta_2 time_t + \beta_3 ln(income_{it}) + \beta_4 ln(income_{it})^2 + \epsilon_{it}$$

This model corresponds to the most traditional EKC specification.[19] As a last benchmark we use the structural identity of Yang and Schneider (1998):

$$carbon_{it} = population_{it} \times \frac{income_{it}}{capita_{it}} \times \frac{energy_{it}}{income_{it}} \times \frac{carbon_{it}}{energy_{it}}$$

The additional data for the last equation is taken form the authors baseline case. We will refer to the 3 benchmarks also as B1, B2 and B3 respectively.

3.5 Model Selection

We apply the described RC methodology to the U.S. carbon emissions state level panel data. The results are summarized in table 2 below. Using the MSFE as our selection criterion,

[17] Up to now they are cited 34 and 73 respectively, according to the Social Science Citation Index

[18] To forecast a time trend Schmalensee et al. (1998) regress the time fixed effects on the year numbers with a two piece spline function that allows for a structural break in the time trend at 1970: $\beta_t = \alpha_0 + \alpha_1 t + \alpha_2(t - 1970) \cdot \mathbf{1}_{\{t \geq 1970\}}$

[19] Holtz-Eakin and Selden (1995) use the last in sample period time fixed effect coefficient as a time dummy in their projections.

the BPM outperforms the benchmarks substantially having a very small $MSFE_{k^*}$ of 65.1 compared to B1 with 610, B2 with 432 and B3 with 458. To provide a scale for those absolute numbers, e.g. benchmark one is off for the year 2000 carbon emissions forecast by a state level average of 16.8%, whereas the BPM is off on average by 8.9%. Those numbers show that there are sizable differences between the benchmarks and the BPM. The BPM forecasting ability is very good compared to the three benchmarks, but by no means significantly better then then 2nd, 3rd or 4th best model in the universe. In fact the models lie close together in their specification form as well as their MSFE. So follows the 2nd best model closely with a MSFE of 66.5. The benchmarks perform comparatively rather poor. Benchmark one is for example outperformed by an overwhelming 30% of all model in the universe. The major source of errors for the benchmark model are their comperativly less good prediction of the per capita emissions for the high pollution states Wyoming, North Dakota and Alaska.

Table 2: Reality Check Results

Tested Benchmarks:	B1	B2	B3	BPM
Test statistic T^{RC}	1540	1038	1111	-4.10
White's RC p-value p_u	0.78	0.79	0.73	1
Critical value at 10%	(4.4e+10)	(4.4e+10)	(4.4e+10)	(14268)
Hansen's RC p-value p_c	0.30	0.29	0.26	1
Critical value at 10%	(2898)	(2071)	(1908)	(65.93)
Hansen's RC p-value p_l	0	0	0	1
Critical value at 10%	(1344)	(889)	(881)	(45.70)
Naive p-value	0	0	0	0.72
Critical value at 10%	(25.15)	(25.24)	(81.28)	(17.59)

Remembering our null hypothesis of no predictive superiority of the best performing model over the benchmark model, we look first at the naive p-value and find it to be zero for all three benchmarks, which is what we would expect given the differences in the MSFE between the models. Intuitively one expects similarly clear results from applying the Bootstrap Reality Check, but as we can see in table 2 this is not the case. The RC p-value centered around zero is surprisingly high with $p_u = 0.78$, $p_u = 0.79$ and $p_u = 0.73$ respectively. Hansen's modified version, using the different centering criteria, yield a p-value of $p_c = 0.30$ and $p_l = 0$ for B1, $p_c = 0.29$ and $p_l = 0$ for B2 and $p_c = 0.26$ and $p_l = 0$ for B3. The large difference between White's and Hanson's RC p-values suggests that in this application poorly performing models have indeed a disproportionate influence on White's test statistic. Further evidence of this can be seen from the critical values in table 2. The critical value using White's centering methodology is unreachable. Even if the best model

was able to perfectly predict, i.e. had zero error, we would not be able to reject the null of no superiority over the benchmark at a 10% significance level.[20] The differences between the p_c and p_l values highlights the influence of some semi-poor models. By simply eliminating the influence of models with $\bar{f}_k$ between $[-A_k, 0]$ on the sample distribution of the test statistic the results turn from no rejection to clear significant rejection of the null. Having over 21 thousand different models, of which a lot are rather bad performing, we do expect poor models to have a large influence on the sample distribution. This is verified by the differences in the different p-values. Therefor it is necessary to use the stricter of the two Hansen centering criteria.

As a verification we also include in table 2 the set of results in which the BPM is used as the benchmark, and the second best model is used as the best alternative. In other words, the null hypothesis for these results is that the BPM has predictive superiority over the second best model. It is reassuring to see that all p-values would suggest that in at least 72% of repeated trials it would be wrong to reject the null. The above results allow us to reject the nulls that each benchmark is not outperformed by the best performing model contained in the given universe of models.

4. Preferred Model Results and Forecasts

We use the best performing model according to the out of sample, which is the one which minimizes the MSFE and is given below:

$$\ln(c_{it}) = \alpha + \rho_i \ln(c_{it-1}) + \ln(income_{it}) + \varepsilon_{it} \tag{6}$$

We estimate the parameters using the whole sample of data, 2091 observations. We explain 99% of the variance in the state level emissions, which is not surprising given that these are time series data and we include lagged emissions. The coefficient estimates and corresponding robust errors of the regression are given in table 3. This simple specification, to the surprise of the authors, has a slightly negative coefficient on income. From a statistical perspective this does not cause an issue, since we are only interested in forecasting performance. From an economic perspective this implies a negative income effect, implying that emissions are an inferior good. Using a more economic approach to model selection, such as the general to specific search, Bayesian shrinkage or the diffusion index approach are

[20] A model with exact CO_2 forecast would have a test statistic $T^{RC} = 1724.50$ in the B1 case and so not even get close to the critical value of 4.4e+10.

likely to result in a different specification. Comparing this model to the literature, Watson et al. (2003) show that a simple AR process dominates most of the models considered in their model universe.

Table 3: Parameter Estimates of the Best Performing Model

Parameter	Estimate	Robust Standard Error
Const.	0.167***	0.019
$ln(income_{it})$	-0.012*	0.006
state specific carbon lags	0.870 – 0.970***	0.005 – 0.014
R^2	0.988	
n	2040	

*** Significant at a 1% level, * Significant at a 10% level

4.1 The Preferred Model and the Kuznets Curve

Before proceeding to apply the BPM to forecasting, it is worth taking a quick look at what it tells us about the existence of a Kuznets type relationship in the full set data. The model does not display the typical inverted-U shape relationship between per capita income and emissions. Per capita income enters the model as a linear function, with the coefficient on income being negative. A potential explanation for the difference between our estimates and those of other studies is the fact that our BPM includes state specific coefficients on the lagged carbon term. This specification is very similar to that obtained by Auffhammer, Carson and Garin-Munoz (2004), who provide an extensive discussion of the information contained in the lag parameter estimates. They interpret the size of the lag coefficients to be indicative of difference in the speed of capital replacement, which varies across provinces/states. The coefficient on lagged carbon emissions tends to be lower for states with higher levels of per capita income. We re-estimated the equation, leaving out the lagged emissions terms and adding a quadratic income term instead. The resulting specification is the classic Kuznets type, containing a constant, an income and an income squared term only. Using this specification the point estimate for the turning point is \$20,204.[21]

[21] As discussed in an earlier version of the paper, one should not place too much emphasis on turning point estimates since it has a large standard deviation, resulting in a huge confidence interval.

4.2 Forecasting US CO_2 Emissions

We now turn to using the BPM to forecast U.S. CO_2 emissions to 2010, which corresponds to the horizon we have used to select our BPM. We compare these forecasts to those in the previous literature and those obtained using the benchmark model. We find that the PBM forecasts are lying in between those obtained using the benchmark models B1/B3 and B2 and are close to the mean prediction of a wide range of studies form the literature.

In order to construct forecasts we require forecasts of the predictors state-level per capita income and population density.[22] We will provide forecasts combining different scenarios for each of two variables population and income. The state population forecasts are based on the projections by Campbell (1996).[23] The income growth scenarios are based on in-sample historical income growth, which are modified to get high, medium and low growth scenarios. Different assumptions about future trends of these explanatory variables are likely to imply very different emissions forecasts. Since it is impossible to include all possible sets of assumptions, we will limit our analysis to combinations of two population projections and three assumptions about the income growth rate.

Unlike for population, there are no official state-level income projections for the U.S. Thus we generate projections based on the projected population growth similar to Auffhammer et al. (2004), controlling for the correlation between population and income. We assume that the income growth rate ξ_t and population growth rate ϕ_t are jointly distributed as $f(\xi_t, \phi_t) \sim N_2(\mu_\xi, \mu_\phi, \sigma_\xi^2, \sigma_\phi^2, \rho)$ and can be characterized in and out of sample by this bivariate normal distribution. The distribution is parameterized by using the in-sample mean and standard deviation of the population growth rate as well as its correlation coefficient with aggregate income growth, μ_ϕ, σ_ϕ and ρ respectively. We consider three different pairs of values for μ_ξ and σ_ξ for constructing the out of sample prediction. These correspond to high, medium, and low income growth scenarios. The medium growth scenario uses the U.S.'s in-sample mean income growth and variance at the state level. For the high growth scenario we raise the average growth by 1% and for the low growth scenario we subtract 1% from the in sample growth mean. We raise/lower the standard deviation by 1.5% for the high and low growth scenario respectively. We calculate $\phi_t \ \forall \ t = 2001, \ldots, 2010$ from Campbell's population projections using the conditional marginal distribution $f_\phi(\xi_t) = N(\mu_\xi - \alpha\mu_\phi + \alpha\phi_t, \sigma_\xi^2(1-\rho)^2)$ where $\alpha = \frac{\rho\sigma_\varsigma\sigma_\phi}{\sigma_\phi^2}$. Using these we obtain state growth rates for personal income. The different

[22] The same procedure that is described below is used for the out-of-sample forecast of the RHS variables in the model selection process, with the only difference that we only use income scenario two and population projections form Wetrogan (1983).

[23] Campbell (1996) provides 2 different estimated population scenarios A and B, whereby he calls A the preferred scenario.

scenario assumptions are summarized in table 4. With the two population projections and the three growth scenarios We have a total of six different population/income scenarios for the emissions forecasting. For the remainder of the paper, we will refer to scenario A2 as the base scenario, since it represents a reasonable mediate path with the preferred population projection A and the medium income growth assumption.

Table 4: Considered growth scenarios for population and income

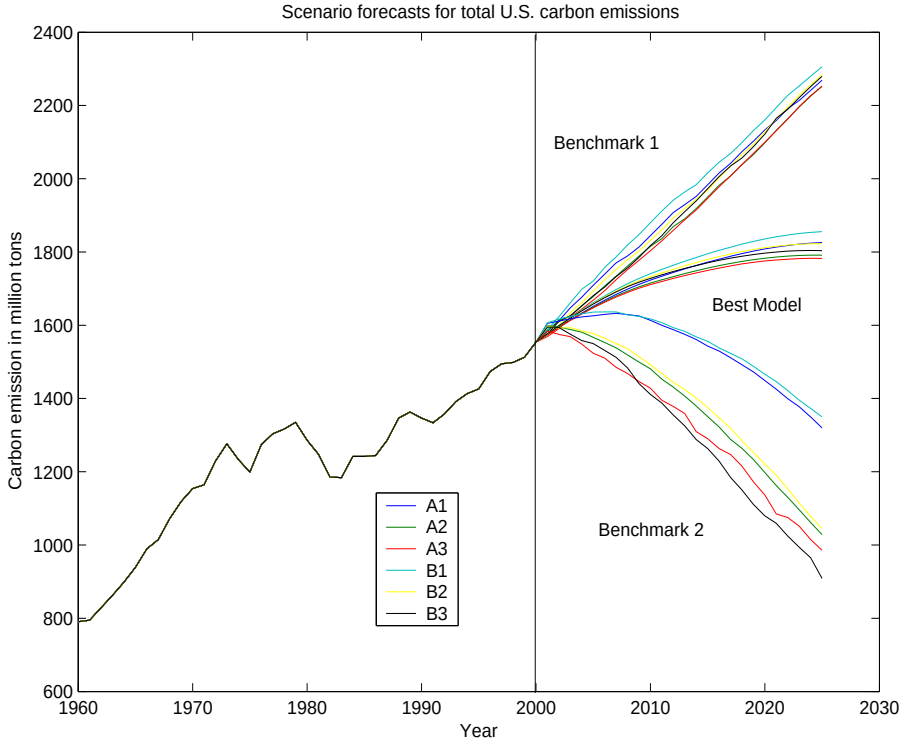
	A1-Slow	A2-Medium	A3-Fast
Population projection	time-series approach		
Income growth average	1.48%	2.48%	3.48%
Income growth std	0.61%	2.11%	3.61%
	B1-Slow	B2-Medium	B3-Fast
Population projection	economic model approach		
Income growth average	1.48%	2.48%	3.48%
Income growth std	0.61%	2.11%	3.61%

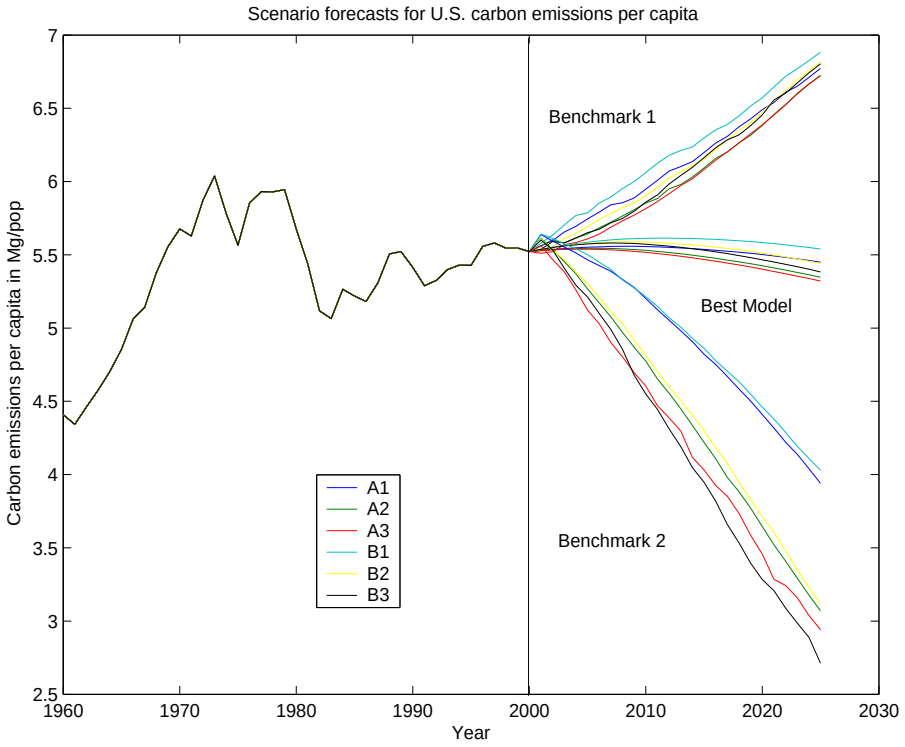
Figure 2 displays the U.S. forecasting results for aggregate CO_2 emissions in terms of carbon output for the BPM and benchmark one. The top graph shows the total emission forecasts for the six different scenarios and the bottom one shows projected per capita emissions. All the scenarios follow by the nature of their derivation a similar shaped trajectory. Per capita emissions for the BPM are decreasing monotonically over the entire forecasting period starting at 2000. Total emissions rise for some time before the go level for all scenarios.

As we see in figure 2 benchmark 1 predicts for all scenarios a much higher carbon emission trajectory than the BPM. Total carbon emissions as well as per capita emissions rise monotonically. Under the base scenario, the benchmark predicts emissions in 2010 of 1.81 billion tons of carbon, which is 6% higher than the BPM predicts under the same scenario. But the divergence is only getting bigger for a longer prediction horizon. Benchmark 2 in comparison predicts much lower total as well as per capita emissions. With 1.48 billion tons for 2010 it predicts 13% lower emissions over this short horizon.

The prediction of benchmark 1 that per capita emissions will rise more steeply than they have at any time since the 1980's seems unlikely given the consistent trend toward lower carbon intensity in the economy. An other problem with benchmark 1's forecast is its poor performs for individual states with high per capita emissions, where we find large discontinuities from the last in sample value to the first prediction. Benchmark 2 on the other hand goes steeply down with its predictions for both total and per capita emissions, which raises doubt about it's accuracy as we observe in the latest EIA reports a slight slowdown

Figure 2: Forecasts of aggregated total carbon emission (top) and per capita carbon emissions (bottom) for the United States of America - Benchmark 1 & 2 versus the Best Performing Model.





but no reversal of the trend for total U.S. emissions.[24] Benchmark 3 has a wide range of predictions for the six scenarios. Scenario A1 and B3 predict a 18% different emissions output for the year 2010 which shows the models sensitivity to changes in population and income growth.

4.3 Comparison with Other Studies

Using the Emission Scenario Database developed for the IPCC's Special Report on Emissions Scenarios, we can readily compare our predictions to those of all previous studies under similar 'no intervention' assumptions. In total there are 26 different, mostly structural studies, with 42 forecasts for the U.S. under scenarios classified as no intervention. Foremost among the previous studies are the IPCC IS92 scenarios, results form the Energy Modeling Forum (Stanford), Nordhaus' DICE model, reports form the Energy Information Administration (U.S. Department of Energy), as well as studies from the Center for Energy and Environmental Policy Research (MIT).

As we can see in the overview in table 5 the different emission estimates of studies for the U.S. for the year 2010 range from 1.30 to 2.36 billion tons of carbon with a mean of 1.77. For the same year we predict 1.73 billion tons using the BPM, and 1.83, 1.51 and 1.90 billion tons of carbon for the three benchmarks respectively .

Table 5: Comparison of forecast results

		Emission Scenario Database	BPM	B1	B2	B3
2010	range	1.30-2.36	1.71-1.74	1.80-1.88	1.41-1.61	1.74-2.06
	mean	1.77	1.73	1.84	1.51	1.90
	std.	0.24	0.01	0.03	0.09	0.12

Overall we see that the BPM provides a forecasts of U.S. emissions that lies very close to the mean of studies in the literature.

5. Conclusions

The current paper tests forecasting performance of a large number of reduced form models against three benchmark models found in the literature. We find that models based on the Kuznets curve as well as a simple version of the IPAT model are outperformed by 15 to 30%

[24]Results are produced by the (Energy Information Agency, 2004) and exist till 2003 so far. They show a drop in total U.S. carbon emissions in 2001 and raising emissions for 2002 and 2003.

of the models considered. Surprisingly the model with the smallest error is a variant of an AR(1) model with a common income effect.

Optimal forecasts for CO_2 emissions serve as important inputs to Global Climate Change models as well as inputs to benefit cost studies. Suboptimal forecasts may have large, consequences for future global climate agreements, since nations with unrealistically high expectations of emissions and therefore costs of emissions reductions may not join such an agreement. The paper shows that future emissions for the United States may be lower than expected based on the benchmark, which is consistent with a recently reported slowdown of aggregate emissions (Energy Information Agency, 2004).

The paper further shows that the Reality Check test provided by White (2000) may be too conservative since it gives too much weight to models with really poor performance. Results suggest that a perfect model may not even beat the benchmark using White's centering. We show that using the corrections provided by Hansen (2004) provides more realistic test results.

While our universe of models includes simple reduced form variants of models based on the IPAT identity (Ehrlich and Holdren, 1971), it would be an interesting future line of research to include forecasts from large scale engineering models, such as the Energy Information Agency's NEMS model, into the universe of models. We are currently pursuing the required forecasts. Further we are working on incorporating forecasts based on aggregate state level as well as US aggregate emissions and test the behavior of the Reality Check with changing horizons. An other alternative would be to use the MSFE on an US level for a model selection criterion.

References

Aldous, D.: 1989, *Probability Approximations Via Poisson Clumping Heuristic*, Springer, New York.

Arrow, K., Bolin, B., Constanza, R., Dasgupta, P., Folke, P., Holling, C., Jansson, B., Levin, S., Mäler, K., Perrings, C. and Pimentel, D.: 1995, Economic growth, carrying capacity and the environment, *Science* **268**, 520–521.

Auffhammer, M., Carson, R. and Garin-Munoz, T.: 2004, Forecasting china's carbon dioxide emissions: A provincial approach, *Working Paper* .

Barbier, E. B.: 1997, Introduction to the Environmental Kuznets Curve Special Issue, *Environment and Development Economics* **2**, 369–381.

Blasing, T., Broniak, C. and Marland, G.: 2004, *Trends: A Compendium of Data on Global Change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN, U.S.A., chapter Estimates of Annual Fossil-Fuel CO2 Emitted for Each State in the U.S.A. and the District of Columbia for Each Year from 1960 through 2001.

Campbell, P.: 1996, Population projections for states by age, sex, race, and hispanic origin: 1995 to 2025, *Technical Report PPL-47*, U.S. Bureau of the Census, Population Division.

Cowles, A.: 1933, Can stock market forecasters forecast?, *Econometrica* **1**, 309–324.

Diebold, F. and Mariano, R.: 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* **13**(3), 253–263.

Ehrlich, P. and Holdren, J.: 1971, Impact of population growth, *Science* **171**(3977), 1212–1217.

Energy Information Agency: 2004, Emissions of Greenhouse Gases in the United States 2003, Website. http://www.eia.doe.gov/oiaf/1605/ggrpt/index.html.

Frey, R. L., Staehelin-Witt, E. and Bloechlinger, H.: 1991, *Mit Oekonomie Zur Okologie: Analyse und Loesungen Des Umweltproblemes Aus Oekonomischer Sicht*, Helbig and Lichtenhahn, Basel and Frankfurt.

Garbaccio, R. F., Ho, M. S. and Jorgenson, D. W.: 1999, Controlling Carbon Emissions in China, *Environment and Development Economics* **4**, 493–518.

Giacomini, R. and White, H.: 2004, Tests of conditional predictive ability, *UCLA mimeo* .

Grossman, G. and Krueger, A.: 1993, *The Mexico-U.S. Free Trade Agreement*, MIT Press, Cambridge, MA and London, chapter Environmental Impacts of a North American Free Trade Agreement, pp. 13–57.

Grossman, G. and Krueger, A. B.: 1995, Economic growth and the environment, *Quarterly Journal of Economics* **110**, 353–377.

Hansen, P. R.: 2004, A test for superior predictive ability, *Working paper* .

Holtz-Eakin, D. and Selden, T. M.: 1995, Stoking the fires? CO_2 emmissions and economic growth, *Journal of Public Economics* **57**, 85–101.

Hoover, K. D. and Perez, T.: 1999, Data mining reconsidered: Encompassing and the general-to-specific approach to specification search, *Econometrics Journal* **2**, 167–191.

IPCC: 2001, *Climate Change 2001, Working Group II: Impacts, Adaptation and Vulnerability*, Summary for Policymakers.

Leamer, E.: 1978, *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, Wiley, New York.

Levinson, A., Harbaugh, B. and Wilson, D.: 2002, Reexamining the empirical evidence for an environmental kuznets curve, *Review of Economics and Statistics* **84**, 541–551.

Lieb, C.: 2005, The environmental kuznets curve and flow versus stock pollution: The neglect of future damages, *Environmental and Resource Economics, volume = 29, number = 4, pages=482-507* .

Lo, A. and MacKinley, C.: 1990, Data snooping biases in tests of financial asset pricing models, *Review of Financial Studies* **3**, 431–468.

Marland, G., Boden, T. and Andres, R.: 2004, *Trends: A Compendium of Data on Global Change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN, U.S.A.,, chapter Global, Regional, and National Fossil-Fuel CO2 Emissions.

McCracken, M. W. and West, K. D.: 2004, Inference about predictive ability, *in* M. P. Clements and D. F. Hendry (eds), *Economic Forecasting*, Blackwell Publishing, pp. 299–321.

Millimet, D., List, J. A. and Stengos, T.: 2003, The environmental kuznets curve: Real progress or misspecified models?, *Review of Economics and Statistics* **85**, 1038–1047.

ONeill, B. C. and Desai, M.: 2005, Accuracy of past projections of us energy consumption, *Energy Policy* **33**(8), 979–993.

Politis, D. N. and Romano, J.: 1994, The stationary bootstrap, *Journal of the American Statistical Association* **40**, 1303–1313.

Schmalensee, R., Stoker, T. M. and Judson, R.: 1998, World carbon dioxide emisions: 1950-2050, *Review of Economics and Statistics* **80**, 15–27.

Selden, T. M. and Song, D.: 1994, Environmental quality and development: Is there a kuznets curve for air pollution emissions, *Journal of Environmental Economics and Mangagement* **27**, 147–162.

Shafik, N.: 1994, Economic development and environmental quality: An econometric analysis, *Oxford Economic Papers* **46**, 757–773.

Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., Kettleborough, J. A., S. Knight, A. M., Murphy, J. M., Piani, C., Sexton, D., Smith, L. A., Spicer, R. A., Thorpe, A. J. and Allen, M. R.: 2005, Uncertainty in predictions of the climate response to rising levels of greenhouse gases, *Nature* **433**, 403–406.

Stock, J. H. and Watson, M. W.: 2002, Macroeconomic forecasting using diffusion indexes, *Journal of Business and Economic Statistics* **20**, 147–162.

Sullivan, R., Timmermann, A. and White, H.: 1999, Data-snooping, technical trading rule performance, and the bootstrap, *Journal of Finance* (5), 1647–1691.

Watson, M. W., Marcellino, M. and Stock, J. H.: 2003, Macroeconomic forecasting in the euro area: Country specific versus area-wide information, *European Economic Review* **47**(1), 1–18.

West, K.: 1996, Asymptotic inference about predictive ability, *Econometrica* **64**(5), 1067–1084.

Wetrogan, S. I.: 1983, Provisional projections of the population of states, by age and sex, 1980 to 2000, *Current population reports.Series P-25* **937**.

White, H.: 2000, A reality check for data snooping, *Econometrica* **68**, 1079–1126.

Yang, C. and Schneider, S.: 1998, Global carbon dioxide emissions scenarios: Sensitivity to social and technological factors in three regions, *Mitigation and Adaptation Strategies for Global Change* **2**(4), 373–404.