International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

The Accuracy of Past Projections of U.S. Energy Consumption

O'Neill, B.C. and Desai, M.

IIASA Interim Report
October 2003

International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

Interim Report **IR-03-053**

The Accuracy of Past Projections of U.S. Energy Consumption

Brian C. O'Neill (oneill@iiasa.ac.at)
Mausami Desai (desai.mausami@epa.gov)

Approved by

Leen Hordijk (hordijk@iiasa.ac.at)
Director

October 27, 2003

Contents

Abstract

Energy forecasts play a key role in the development of energy and environmental policy. Evaluations of the accuracy of past projections can provide insight into the uncertainty that may be associated with current forecasts. They can also be used to identify sources of inaccuracies, and potentially lead to improvements in projections over time. Here we assess the accuracy of projections of U.S. energy consumption produced by the Energy Information Administration over the period 1982-2000. We find that energy consumption projections have tended to underestimate future consumption. Projections 10 to 13 years into the future have had an average error of about 4%, and about half that for shorter time horizons. These errors mask much larger, offsetting errors in the projection of gross domestic product (GDP) and energy intensity. GDP projections have consistently been too high, and energy intensity projections consistently too low, by more than 15% for projections of 10 years or more. Further work on the source of these sizable inaccuracies should be a high priority. Finally, we find no evidence of improvement in projections of consumption, GDP, or energy intensity since 1982.

Acknowledgments

This paper has been submitted to *Energy Policy* for publication.

We thank Nico Keilman and an anonymous reviewer for helpful comments. Several Energy Information Administration staff members kindly provided guidance on Annual Energy Outlook methodology and results: Mary Hutzler, Paul Holtberg, Tom Peterstik, Eugene Reiser, and Esmerelda Sanchez. We also thank Katherine Fisher and Ruth Siegel for assistance with early stages of this paper.

About the Authors

Brian C. O'Neill is a Research Scholar at IIASA and Assistant Professor (Research) at the Watson Institute for International Studies, Brown University, Providence, RI, USA.

Mausami Desai attended the IIASA Young Scientists Summer Program (YSSP) in the summer of 2003. She is currently at the Clean Air Markets Division, U.S. Environmental Protection Agency, Washington, D.C., USA. This work was completed while she was a Research Assistant at the Watson Institute for International Studies, Brown University, Providence, RI 02912, USA.

The Accuracy of Past Projections of U.S. Energy Consumption

Brian C. O'Neill and Mausami Desai

Introduction

Projections of future energy consumption play a key role in many analyses of energy and environment. They serve as a basis for planning within the energy industry; for research questions regarding future energy production, consumption, and environmental impacts such as air pollution or climate change; and for evaluating the need for and potential effects of energy and environmental policies. A noteworthy example of a current policy application is the evaluation of proposed climate change policies. For example, estimates of the costs of meeting the commitments agreed to in the Kyoto Protocol range widely (Weyant 2001) and depend strongly on assumptions about what emissions would be in the absence of policy, which in turn rely heavily on projected energy use. Similarly, U.S. climate policy announced in 2002 by the Bush Administration calls for an 18% reduction in the carbon intensity of economic production within the U.S. by 2012 (Bush Administration 2002). Evaluating what kind of action might be necessary to achieve this goal requires a projection of energy intensity and fuel mix in the absence of such a policy.

Analysis of the performance of past projections can be instructive for two main reasons. First, it can provide useful information for characterizing the uncertainty in current projections. The typical magnitude of errors in past projections over a given projection horizon can serve as a guide in quantifying uncertainty over similar time horizons into the future. Of course there is no guarantee that the performance of current projections will be similar to the performance of past ones, and other sources of information should be, and generally are, used to quantify uncertainties. Nonetheless, historical error analysis can be useful as a benchmark. A second motivation for analyzing past projections is the possibility of improving current projections. By identifying sources of error, research can be focused on improving the components of projections that would have the largest payoff in terms of improved outcomes.

There are many examples of useful analyses of forecast errors in the energy field and in other fields as well, such as agriculture (McCalla and Revoredo 2001) and population. For example, errors in population projections have been extensively analyzed (Keilman 1999), and a recent NAS report (National Research Council 2000) encourages the use of such analysis in informing judgments on uncertainty in current projections. The most recent probabilistic projections of global population produced by IIASA (Lutz et al. 2001) operationalize this recommendation. The authors used the results of historical error analysis in UN population projections to define lower bounds to uncertainty in projected population size. Error analysis has also identified key sources

of errors in projected population size. Bulatao (2001) shows that errors in baseline data – the estimated population size, age structure, fertility, and mortality in the starting year of the projection – play a key role. Thus substantial improvements in forecasts would be possible by improving the quality and coverage of the data on which projections are based.

Several studies have analyzed the performance of energy forecasts for the U.S. or for the world. Smil (2000) catalogues the many failures of long-term energy forecasts, which he describes as having "missed every important shift of the past two generations," (p. 262) including the first and second oil crises, the post-1970 reduction in electricity demand in industrialized countries, and the cumulative contributions of energy conservation. Based on this review, he recommends that efforts to forecast the energy system be replaced by the development of scenarios designed to explore alternative futures rather than to predict the most likely one (see, e.g., Silberglitt et al. 2003). Other studies focus on quantitative analyses of errors in particular modeling efforts. For example, the performance of projections of oil prices is a particularly well-known example of the hazards of forecasting. Following the 1979-1980 oil price increases, most analysts expected steadily rising oil prices. In fact, nominal prices fell to less than 50% of their 1981 value by 1986 (Schrattenholzer 1998), and projections of oil prices for the year 1990, made in 1980 by a number of models for the Energy Modeling Forum, were off by a factor of 2 to 3 (Huntington 1994). These projections exhibit what demographers (Keilman 1999) and others call "assumption drag," or the tendency for forecasters to be slow to incorporate new information (e.g., changing oil market conditions in the case of energy, widespread declines in fertility in the case of population) into their forecasts.

Error analysis for energy forecasting has also yielded insight into sources of error. Huntington (1994) found that the sources of errors in oil price forecasts over the 1980s varied with the time horizon of the projection. Inaccuracies over the first half of the decade were driven by model inputs, particularly inaccurate projections of gross domestic product (GDP) and of expansion of non-OPEC oil supply. Inaccuracies in the late 1980s were due mainly to inadequate demand responses to price changes in the models. The analysis also notes that in contrast to the extremely inaccurate price forecasts, concomitant projections of consumption were inaccurate by only about 2%, demonstrating the price inelasticity of much of the demand for oil. Linderoth (2002) examines errors in forecasts of energy consumption made by the International Energy Agency of OECD countries over the period 1978-1994. He concludes that inaccurate GDP projections have been significant contributors to these errors, and that energy price changes have also played a significant role.

A handful of studies have focused on projections of the U.S. energy system. Craig et al. (2002) analyze projections made before 1980, concluding that forecasters underestimated the importance of surprises such as the oil embargoes of the 1970s and the subsequent increase in energy efficiency. Projections of consumption in the year 2000 were uniformly too high. Cohen et al. (1995) analyze projections by the U.S. Energy Information Administration (EIA) made between 1978 and 1993, finding, as in other analyses, that price forecasts have been far less accurate than projections of production or consumption. In addition, they find that projections greatly improved between 1978, a time strongly affected by the oil crises, and the early 1980s, by which time the effect of these disruptions had begun to dissipate. Most large errors in early

forecasts were due to combinations of both high price assumptions (based on then current experience with oil crises) and the assumption that regulations then in place would remain so, when in fact they were often drastically modified or repealed. Shlyakhter et al. (1994) analyzed projections by the U.S. Energy Information Administration made between 1983 and 1987 of 1990 U.S. energy production and consumption by sector in order to derive distributions of errors. They then used those distributions to specify uncertainty intervals for current forecasts. They find that commonly assumed normal distributions of errors substantially underestimate the frequency of extreme outcomes in historical experience.

Since 1996, the EIA itself has analyzed the performance of its own projections (e.g., Holte 2001). Their analyses find that changes in energy policies have had a major impact on forecast accuracy, that price forecasts have continued to be less accurate than forecasts of production or consumption (and have typically been too high), and that projections related to natural gas have been less accurate than those related to other fuels. Their methodology is to calculate mean absolute percent errors in EIA forecasts for various quantities, and to average these errors over all projections and all time horizons.

In this paper we analyze the EIA medium-term projections of U.S. energy consumption. The EIA has published these projections in its Annual Energy Outlook (AEO) each year since 1982. Because they were produced within a single institutional setting, with a relatively stable methodology (discussed in more detail below), the AEOs provide a meaningful basis for error analysis. The 20-year history of projections, and time horizons for individual projections ranging from 8 to 24 years, provide a sufficient basis for determining indicators of average performance as a function of the time horizon of a projection. The AEOs typically contain a reference projection, as well as several variants. Our aim is to analyze the accuracy of the reference projections of total consumption, identify the major sources of error, and look for evidence of improvement in accuracy over time.

Our analysis differs from previous work in several ways. First, we focus on quantifying errors and their sources. In contrast, the aim of Shlyakhter et al. (1994) is to quantify the degree of overconfidence in uncertainty ranges assigned by forecasters in past projections. In examining sources of errors, we differentiate between contributions from errors in baseline data, errors due to cyclical (or inter-annual) variability in consumption that projection models are not designed to forecast, and errors in projecting the trend in energy consumption, important distinctions that have not been made in previous work. We also decompose errors into contributions from errors in GDP growth and errors in energy intensity. While Cohen et al. (1995) also distinguish between these two components, they use a more limited dataset and do not correct for baseline and variability errors. We explicitly control for time horizon in assessing projections. In contrast, the EIA's analysis of its own projections (Holte 2001) averages errors for a given projection across time horizons, making it difficult to compare different projections that may have spanned different lengths of time or have different availability of output data, and obscuring patterns in accuracy over different horizons. Finally, the focus of our analysis is on projections made since 1982, a period which has been uninterrupted by major crises in the global or national energy system. In contrast, other studies (Craig et al. 2002; Cohen et al. 1995) include analysis of long-term projections made before or during the crisis periods of the 1970s, and therefore their

primary conclusions are based on the performance of models with respect to these special conditions.

In the next section we describe the models used by EIA and the data (model inputs and projection results) on which the analysis is based. Following that, we define the measures of error we employ. We then report our results and conclude with discussion and directions for future work.

Models and Data

EIA Forecast Models: IFFS and NEMS

The U.S. Department of Energy has been producing energy projections since 1974. However, available output from projections before 1982 is insufficient to add meaningfully to our analysis. Between 1982 and 1993, EIA projections were produced using the Intermediate Future Forecasting System (IFFS) model. The IFFS is an engineering-economic model of all U.S. energy markets. It can be considered a partial-equilibrium framework that focuses on energy and excludes other, non-energy goods and services produced in the U.S. economy. It represents the U.S. energy system using four end-use demand modules (residential, commercial, transportation, industrial), two supply modules (oil and gas, coal), and two conversion modules (electricity, petroleum refining). Regional disaggregation varies by module but is typically at the level of 10 federal regions. In addition, a macroeconomic module allows for feedback between domestic macroeconomic indicators such as GDP, and world energy prices. Macroeconomic growth paths are determined beginning with an exogenous growth case taken from simulations by Data Resources, Inc. (DRI), which is now part of Global Insight, Inc. The module then calculates an adjusted growth path that takes into account feedbacks from the energy system by iteratively calculating first demand and prices based on the macroeconomic indicators, and then using a reduced-form representation of DRI models to estimate the influence of energy price changes on the macroeconomy. The world crude oil price is taken to be exogenous, derived from a separate EIA projection using a global model.

The National Energy Modeling System (NEMS) replaced IFFS beginning with the 1994 AEO. The basic structure of NEMS is similar to IFFS (Energy Information Administration 1994, 2003a). It generally operates at the level of nine regions within the U.S. (census divisions), and one non-U.S. region. NEMS breaks down the energy system into the same demand and conversion sectors as in IFFS, but adds two additional supply modules for a total of four (oil and gas, renewables, natural gas transmission and distribution, coal). The main difference from IFFS is that many of these modules explicitly represent individual technologies (others, e.g., industrial demand and oil and gas supply, use more limited representations), and that provision is made for technological improvement over time. The expanded capabilities of NEMS were motivated by the technology policy and regulatory issues that had arisen in the early 1990s, such as improvements to the Clean Air Act Amendments of 1990, restructuring of electricity markets, and the integration of renewable technologies (Mary Hutzler, personal communication). Similarly to IFFS, a macroeconomic module allows for feedback between domestic macroeconomic indicators and energy prices. However,

NEMS also allows for feedback between world oil prices and energy supply/demand in the U.S. An international module assumes a reference non-U.S. oil supply and demand, and then calculates a world average oil price based on the assumption that marginal changes in non-U.S. production come from OPEC. Thus the world price is determined in each time step by forecasting based on the price in the previous time step and the percent utilization of OPEC production capacity. In turn, the new price, through its affect on demand and oil imports, affects OPEC production capacity.

The EIA uses NEMS to produce a reference forecast, and additionally four variants assuming higher or lower economic growth, or higher or lower world oil prices. In addition, it produces a large number of special variants testing sensitivities to individual assumptions or policies. Our analysis is based on the reference forecast in each AEO.

Data

The data set for our analysis consists of past projections of U.S. energy consumption, GDP, and energy intensity (EI) contained in the 1982-2002 AEOs[1] (Energy Information Administration 1983-2003). The time horizon for projections varies with the projection; for example, the 1982 AEO makes projections only until 1990, whereas the 1998 AEO projects to 2020. The basic dataset is available in EIA's 2001 *Annual Energy Outlook Forecast Evaluation* (Holte 2001), but this source does not include values for all years in all projections. We include additional data obtained directly from the original AEO publications (Energy Information Administration 1983-2003).[2] Consumption data from AEOs published before 1990 are adjusted to include consumption of energy from dispersed renewables, to be consistent with later projections (see Appendix for details).

Actual values for total energy consumption through 2001 were taken from the *Annual Energy Review 2001* (Energy Information Administration 2002), and a value for 2002 was taken from the most recent available *Monthly Energy Review* (Energy Information Administration 2003b). EIA compiles estimates of the actual quantities of aggregate consumption by summing consumption in the residential, commercial, industrial, transportation, and electric power sectors. Within each sector, EIA collects consumption data by fuel type from suppliers through required surveys.

The real GDP projections were taken from the original AEOs, which report projected real GDP (in units of U.S. dollars expressed in terms of a particular base year) along with projected implicit price deflators. In order to compare real GDP projections across AEOs, all projected real GDP values were converted from various base years to 1996 dollars using actual 1996 chain weighted implicit price deflators as reported by the Bureau of Economic Analysis (2003):

---

[1] An AEO does not exist for 1988 because of a change in the naming convention. The 1989 AEO follows the 1987 AEO.

[2] In some instances, the data within the AEOs is presented at 5-year intervals, rather than annually for a set time horizon. The EIA models produce annual projection values, though they are not always published in the AEOs. AEOs 1996 thru 2003 are available online at EIA's website http://www.eia.doe.gov/oiaf/archive.html. Previous AEO publications (1982-1995) are available in print or microfiche.

$$GNP\ or\ GDP_{real,baseyear=1996} = \left(GNP\ or\ GDP_{real,baseyear=yyyy}\right) * \frac{Deflator_{yyyy}}{Deflator_{1996}} \quad (1)$$

Estimates of actual real GDP in chain-weighted 1996 dollars were obtained from the AER 2001 (Energy Information Administration 1983-2002 (2001, p. 353)) and are consistent with those published by the Bureau of Economic Analysis (2003). For AEOs 1982 through 1992, economic growth was measured in terms of gross national product (GNP), rather than GDP. For consistency, the actual historical nominal values for GNP in those years were obtained from the Bureau of Economic Analysis (2003). Projected and actual values of energy intensity are simply the ratio of energy consumption to GNP or GDP for a given year.

Error Definitions and Decomposition Methodology

We examine four error types using several different measures of error. In defining them, it is useful to first distinguish among various measures of time:

t	calendar year being projected
τ	projection year
t_b	base year of the projection
$TH = t - t_b$	time horizon

The base year, t_b, is the most recent year in which consumption is estimated from data rather than projected with the model. It is often, but not always, the same as the projection year, τ, which indicates the year in which the projection was made. For example, for the 1989 AEO, the base year was 1988, so in that case τ = 1989 but t_b = 1988. The time horizon, TH, indicates the length of the projection. The projection for 1995 from the 1989 AEO has a time horizon of 7 years (1995-1988), while the projection for 2000 has a time horizon of 12 years.

The first type of error we focus on is *visible error* (*V*), which indicates the difference between the projected energy consumption and actual (observed) consumption for a given year, or

$$V_\tau(t) = \hat{E}_\tau(t) - E(t) \quad (2)$$

where the subscript τ is the projection year, $\hat{E}$ is projected energy consumption and *E* is actual energy consumption. Visible error is most relevant to users of projections who want to know how accurate they are; it reflects the error the user actually "sees" in the projection. But to understand the source of the visible error, it is useful to examine "invisible" errors; i.e., components of the visible error whose combined effect produces the net visible error (Bulatao 2001). Here we decompose visible error in consumption in two different ways. First, we decompose it into three components of error in consumption: *baseline error (B)*, *trend error (T)* and *variability (Var)*, so that

$$V_\tau(t) = B_\tau + T_\tau(t) - Var(t) \quad (3)$$

Figure 1 illustrates the relationship among these types of error. The baseline error captures errors in the initial estimates for energy consumption in the base year and is calculated as

$$B_\tau = \hat{E}_\tau(t_b) - E(t_b) \qquad (4)$$

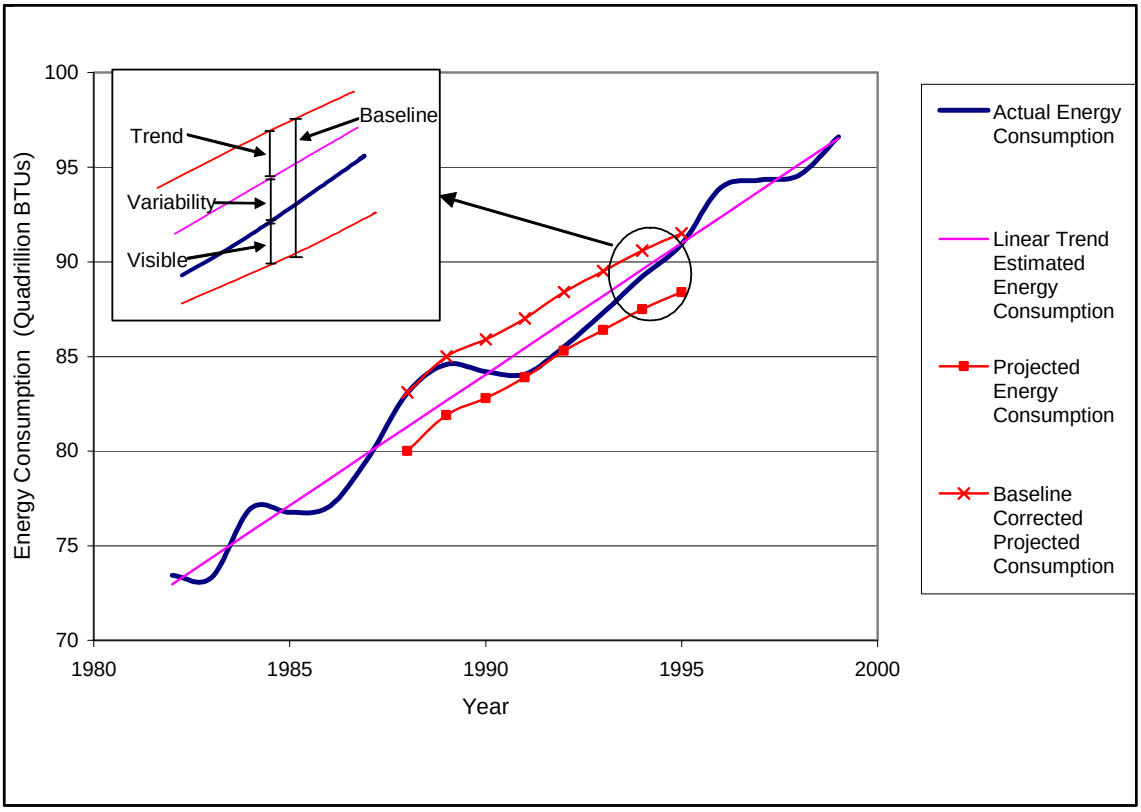

Figure 1. Types of error.

Projections that begin with inaccurate estimates of consumption in the base year are likely to project future consumption inaccurately even if the model is otherwise very accurate. To account for this, we assume that the baseline error is constant across all time horizons for a given projection. For example, if the baseline error is +X Btu (that is, the projection overestimates consumption in the base year), then we assume that consumption is overprojected by X Btu in all future years. This is a simplification; in principle, it would be necessary to rerun the projection model with corrected baseline data but with assumptions and parameter values otherwise identical to the original projection in order to calculate the effect of the baseline error on the projection over time. Since for practical reasons that is impossible, we adopt the constant baseline error assumption as a reasonable simplification.

The *trend error* measures the deviation of the projection (corrected for baseline error) from the historical trend. The rationale for this kind of error is that the models used to project consumption are designed to project longer-term trends in consumption, not interannual variability. A large visible error in a given year could be generated from a short-term fluctuation in consumption due to fluctuations in oil prices, weather, or other factors, which the models do not attempt to predict, and therefore in those cases the visible error would not be indicative of the model's performance. To control for this possibility, we define a trend in historical energy consumption, $E_T(t)$, as a linear fit to

annual consumption data (assuming a nonlinear trend using polynomials had little effect on results, since the 1982-2000 period saw a roughly linear increase in consumption). Trend error is calculated as

$$T_\tau(t) = \hat{E}_\tau(t) - B_\tau - E_T(t) \quad (5)$$

Finally, *variability error (Var*) measures the deviation of actual consumption from the historical trend,

$$Var(t) = E(t) - E_T(t) \quad (6)$$

Note that the variability error is independent of the projection: it is determined only by the difference between actual consumption and the historical trend.

The second way in which we decompose visible error is by expressing it as a sum of errors in the forecasts of GDP and of energy intensity (EI, the ratio of energy consumption to GDP). Since GDP forecasts are essentially (although not completely) exogenous to the EIA consumption forecasts, it is worth examining the degree to which errors in consumption are due to errors in forecasts of GDP, or to errors in forecasting energy intensity (i.e., energy consumption given a particular forecast of GDP). While the two components are not entirely independent, since macroeconomic assumptions affect the forecast of energy intensity, a comparison of their errors can still be informative. We report the results of this analysis later.

For both decompositions, we use different measures to analyze each type of error.

- The percentage error (PE) measures the proportional error at a particular point in time and provides a sense of both the degree and direction of the error.
- The absolute percentage error (APE) is used to assess the magnitude of point errors independent of their direction.
- The mean percentage error (MPE) can be a useful indicator of bias, or the tendency to over- or under-predict consumption. Because the MPE averages over a signed quantity (i.e., PE), it is affected by canceling. Positive errors will cancel negative ones, yielding a small mean error even if individual point errors are large. Thus, it is not a good indicator of accuracy, but still yields useful information on bias. The MPE can be calculated over time within a given projection, or across projections for a given time horizon.
- The mean absolute percentage error (MAPE) controls for the offsetting of negative and positive percentage errors and is therefore more informative about the average magnitude of the errors, independent of sign.

Results

Energy Consumption

Figure 2 shows the full series of AEO projections of total consumption, compared to the observed values. Observed values are taken to be the historical estimates reported in the AER 2001. Considered as a group, the projections appear to have tended to underestimate actual consumption.

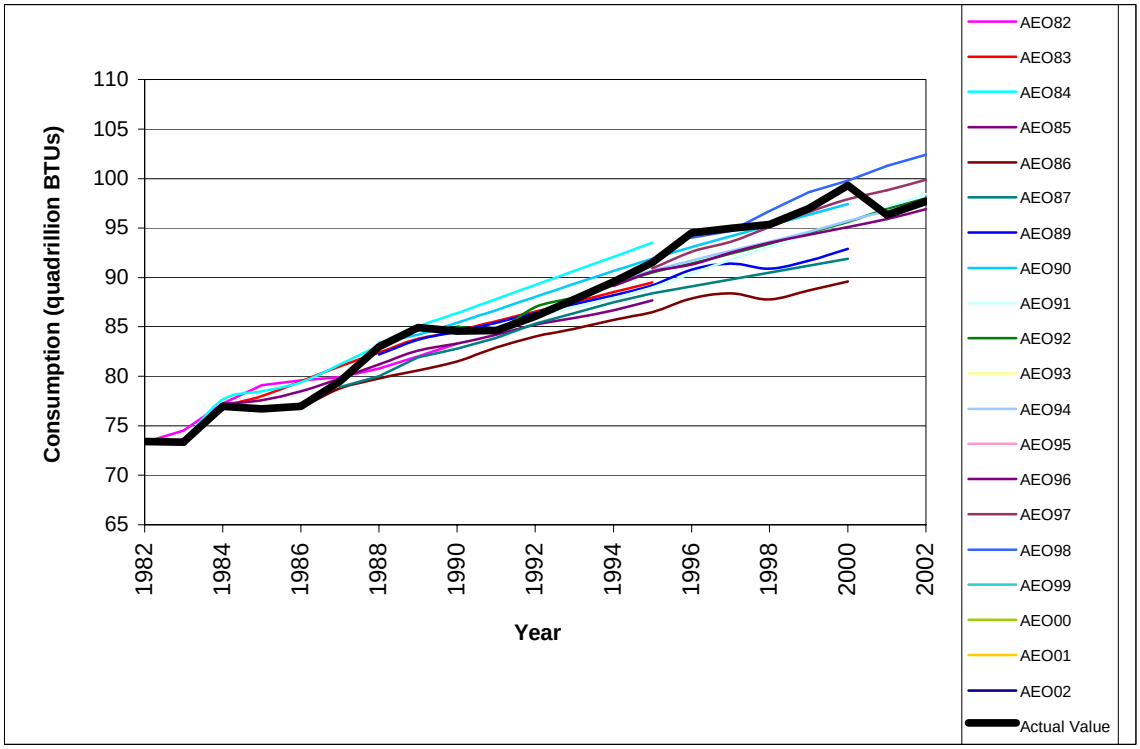

Figure 2. Projections and actual values.

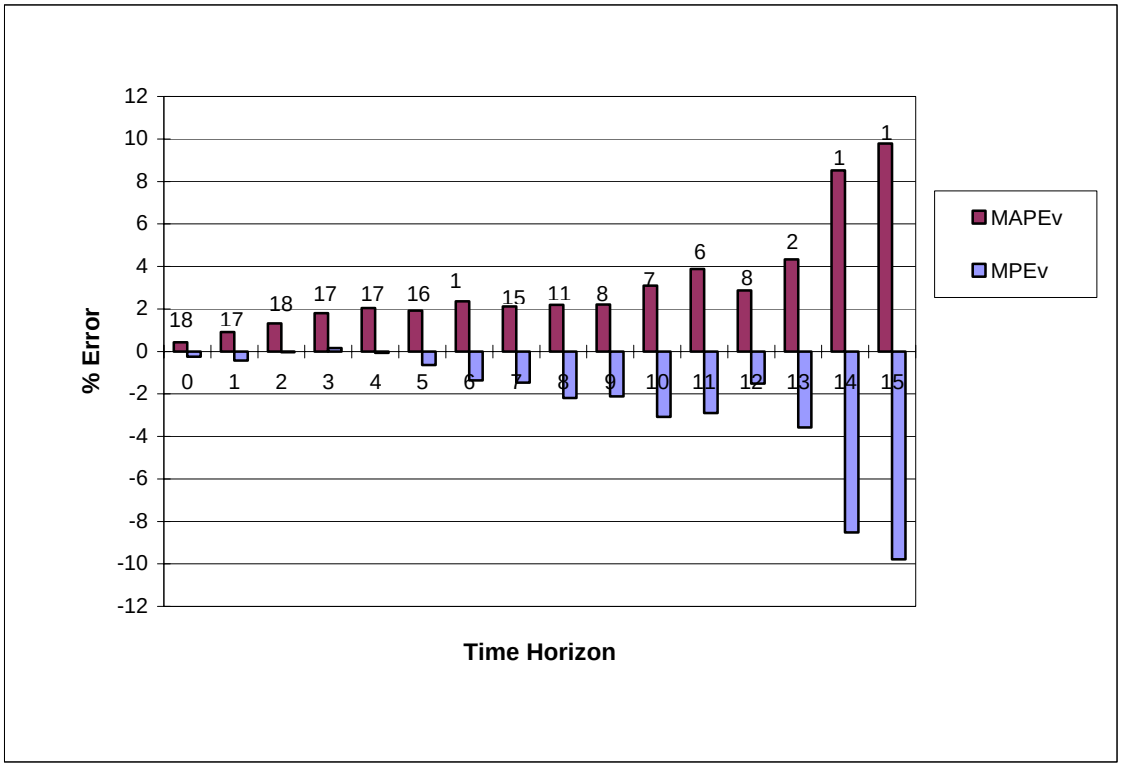

Figure 3. Mean percentage visible error, mean absolute percentage visible error in projected energy consumption by time horizon, and sample size.

Figure 3 shows the mean percentage visible error (MPE_V) by time horizon. Recall that the MPE_V is an average of the signed values of the percentage errors across projections, which allows for canceling of positive and negative errors in individual projections. It therefore reflects whether projections in general have tended to be high (positive MPE_V) or low (negative MPE_V) at particular time horizons. The figure confirms that consumption projections have indeed tended to be low over time horizons of five years or more; at shorter time horizons, the bias in the projections is essentially zero. Figure 3 also shows the mean *absolute* percentage visible errors ($MAPE_V$) in total energy consumption by time horizon. This measure indicates how accurate, on average, the EIA projections have been, independent of whether they have been high or low, controlling for time horizon. The mean absolute visible error is quite low – about 2% – for time horizons up to nine years, and grows slightly to 3-4% for 10- to 13-year projections and to 8-10% at a 14- to 15-year time horizon. However, sample size (indicated in the figure as labels on top of each bar) drops sharply with increasing time horizon; results for 14- and 15-year horizons are based on only a single projection and cannot therefore be considered a reliable measure of forecast accuracy.

It is worth examining the contributions of component errors to both $MAPE_V$ (as an indicator of projection accuracy) and MPE_V (as an indicator of projection bias). Figure 4a shows a decomposition of $MAPE_V$ into baseline error, variability, and trend error. It shows that baseline error makes a relatively small contribution to the inaccuracy of the consumption projections. Inaccuracy in projections is mainly attributable to variability and trend error. Over time horizons of about seven years or less, variability and trend error are each responsible for about half the visible error in projections (i.e., each contributes about one percentage point to the visible error of about 2%). At longer time horizons, trend error grows to two to three times the size of variability. These results are significant for at least two reasons. First, variability error gives an indication of the irreducible component of the error in consumption projections. Since the projection model is not designed to forecast inter-annual variability, roughly 1% error (on average) in projections is likely unavoidable, even if baseline and trend error were reduced to zero. Second, the mean absolute error in the trend ($MAPE_T$) is the measure that is most relevant to evaluating the use of the projection model itself, since it corrects for baseline error and is also unaffected by variability error. Neither of these error types reflects inaccuracies generated by the projection model. Thus it is $MAPE_T$ that is most useful in, for example, diagnosing sources of inaccuracies within the IFFS and NEMS modeling systems. The figure shows that, on average, $MAPE_T$ behaves similarly to the $MAPE_V$, remaining quite low for time horizons of about seven years or less, and increasing at longer time horizons.

Figure 4b, which decomposes MPE_V, gives a sense of the source of the bias toward under-prediction in longer-term projections. It demonstrates that this bias is primarily the result of under-prediction of the trend, since it is only MPE_T that shows increasingly negative values for longer time horizons. MPE_{Var} is very small, and MPE_B is essentially zero, at all time horizons, since baseline and variability errors show no systematic bias and tend to cancel across projections. Thus, while baseline and, especially, variability errors are partly responsible for the inaccuracies in the EIA projections, it is trend error alone – a direct reflection of the performance of the projection model – that produces the bias toward under-prediction of consumption.

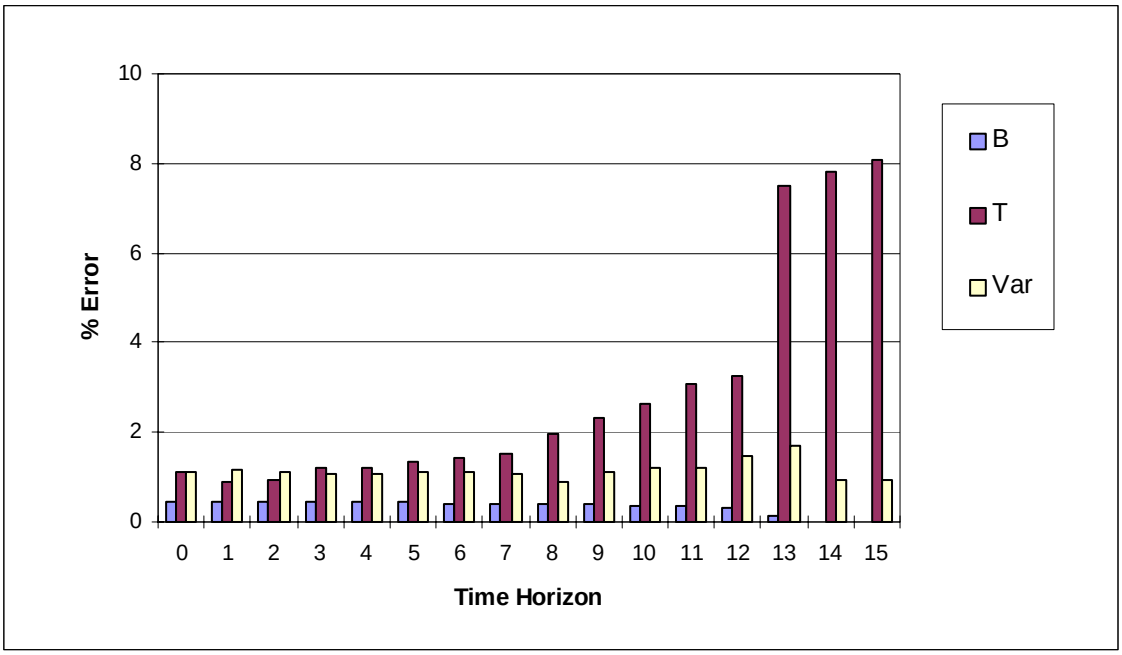

Figure 4a. Error decomposition: Absolute percentage error in projected energy consumption by time horizon.

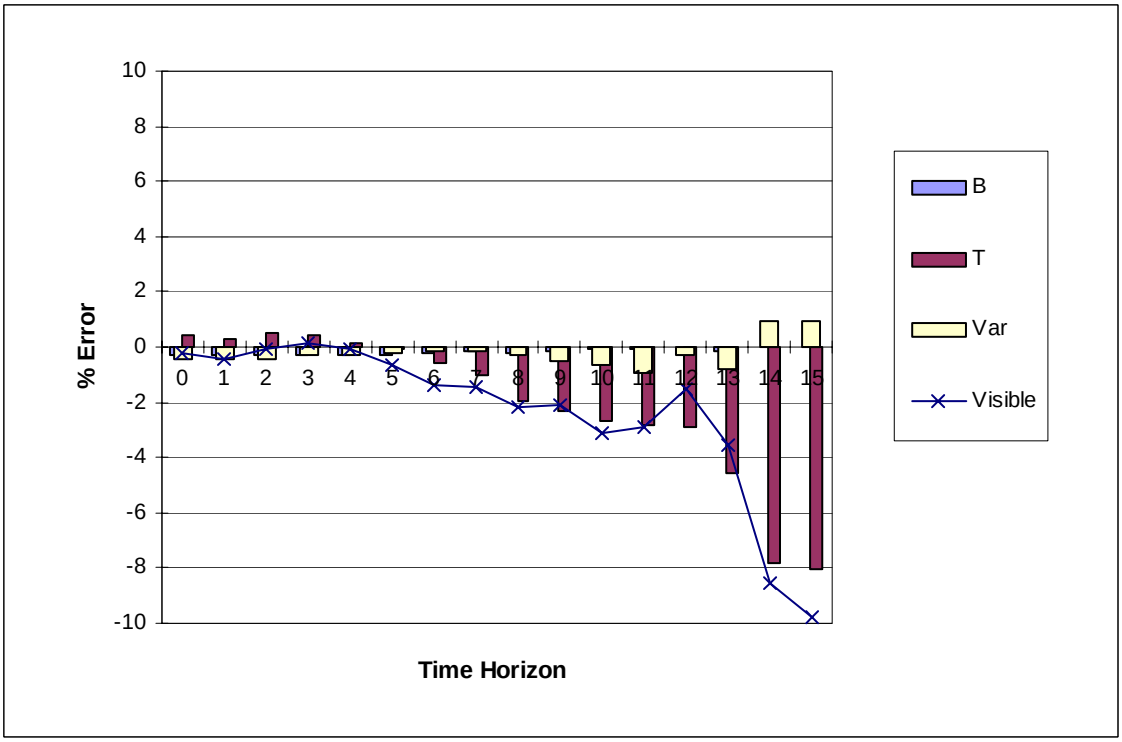

Figure 4b. Error decomposition: Mean percentage error in projected energy consumption by time horizon.

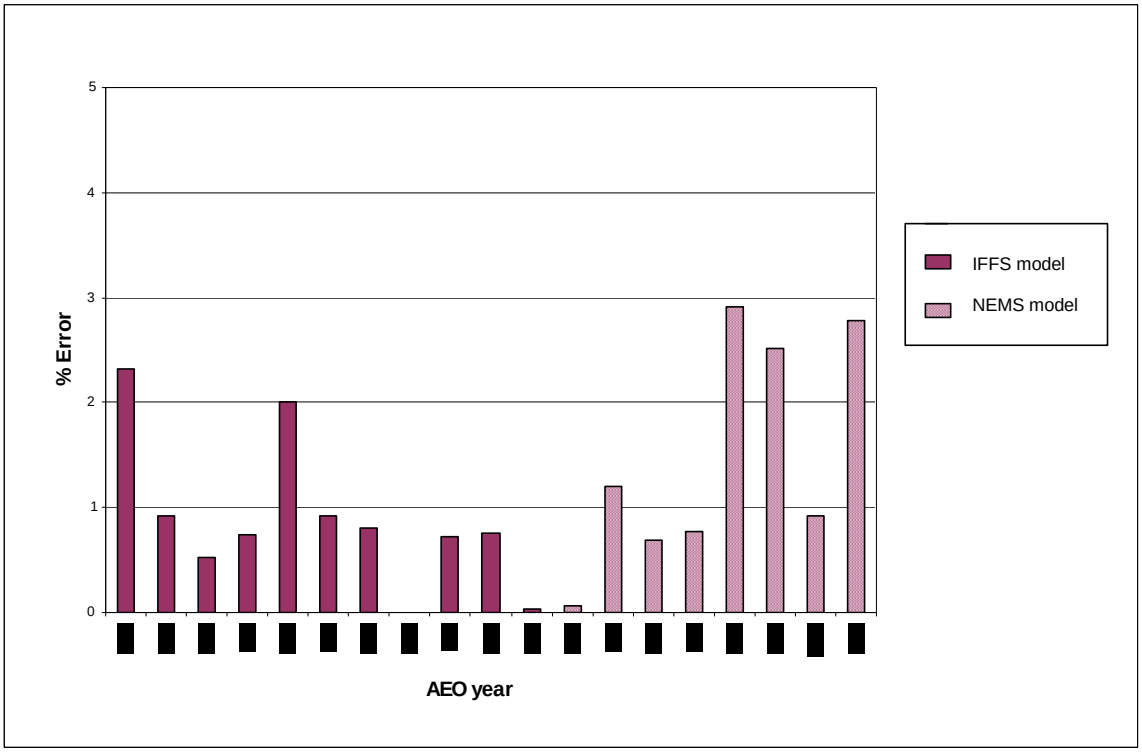

Figure 5a. Mean absolute trend error: Three-year projections.

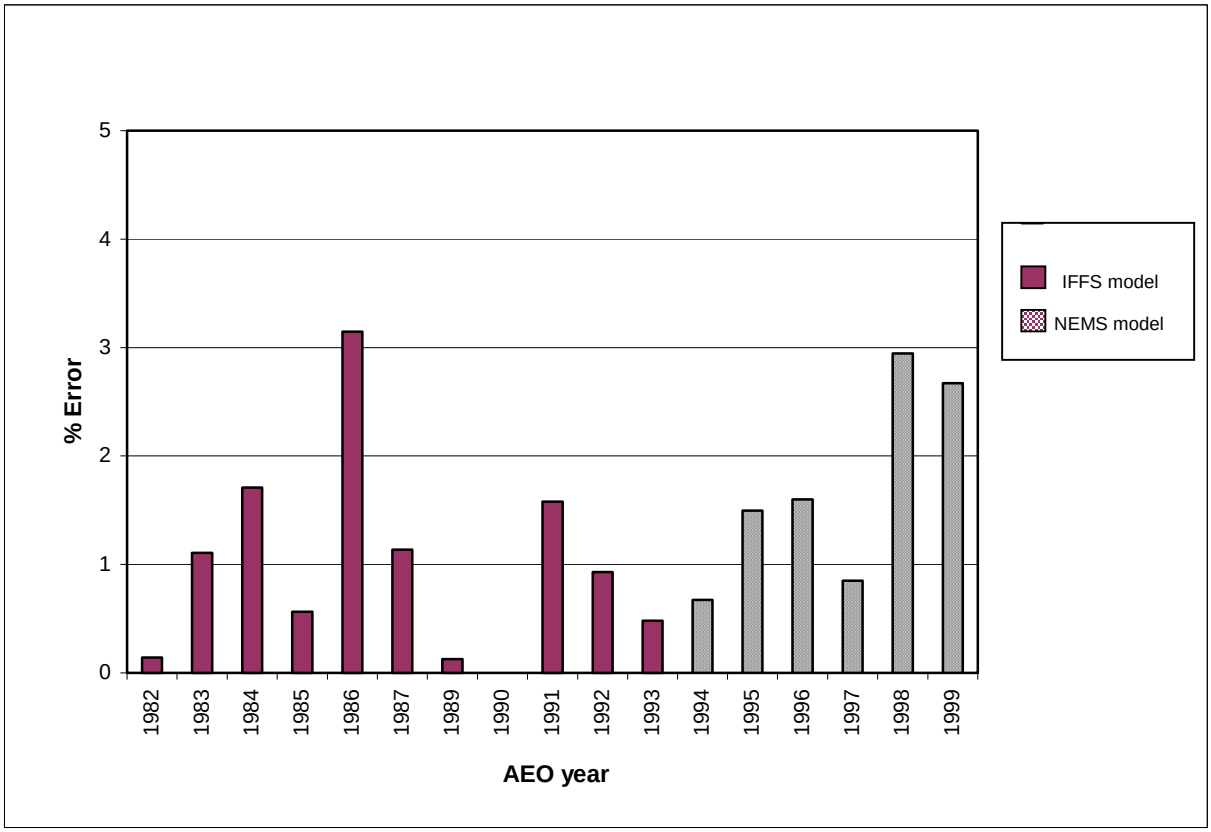

Figure 5b. Mean absolute trend error: Five-year projections.

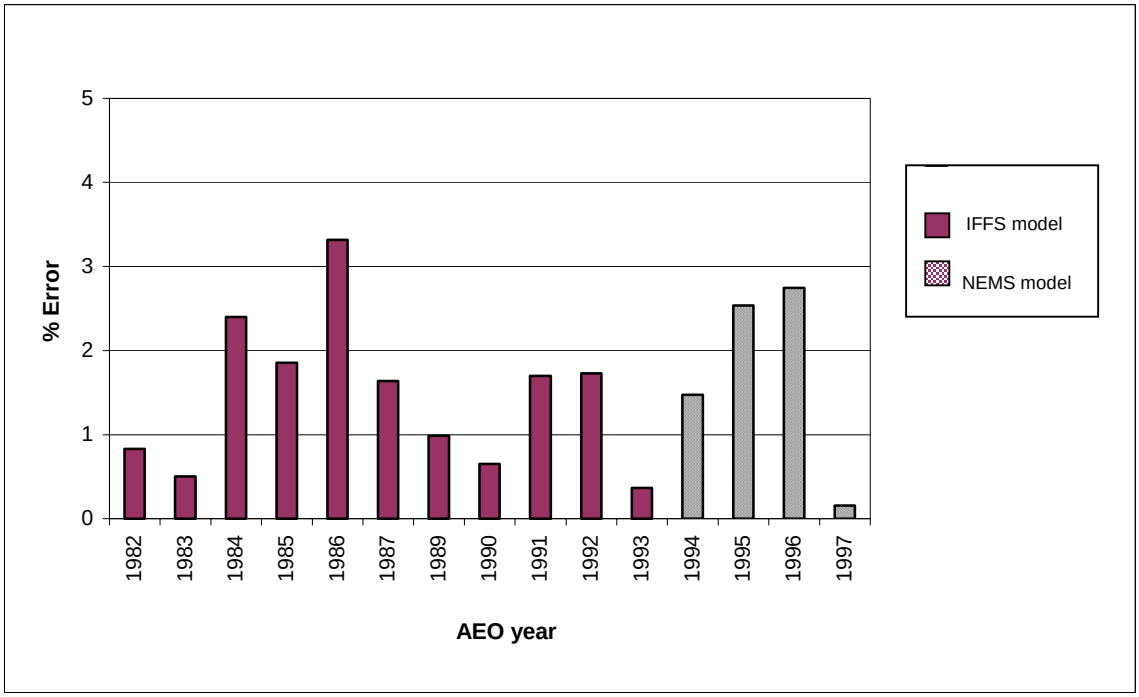

Figure 5c. Mean absolute trend error: Seven-year projections.

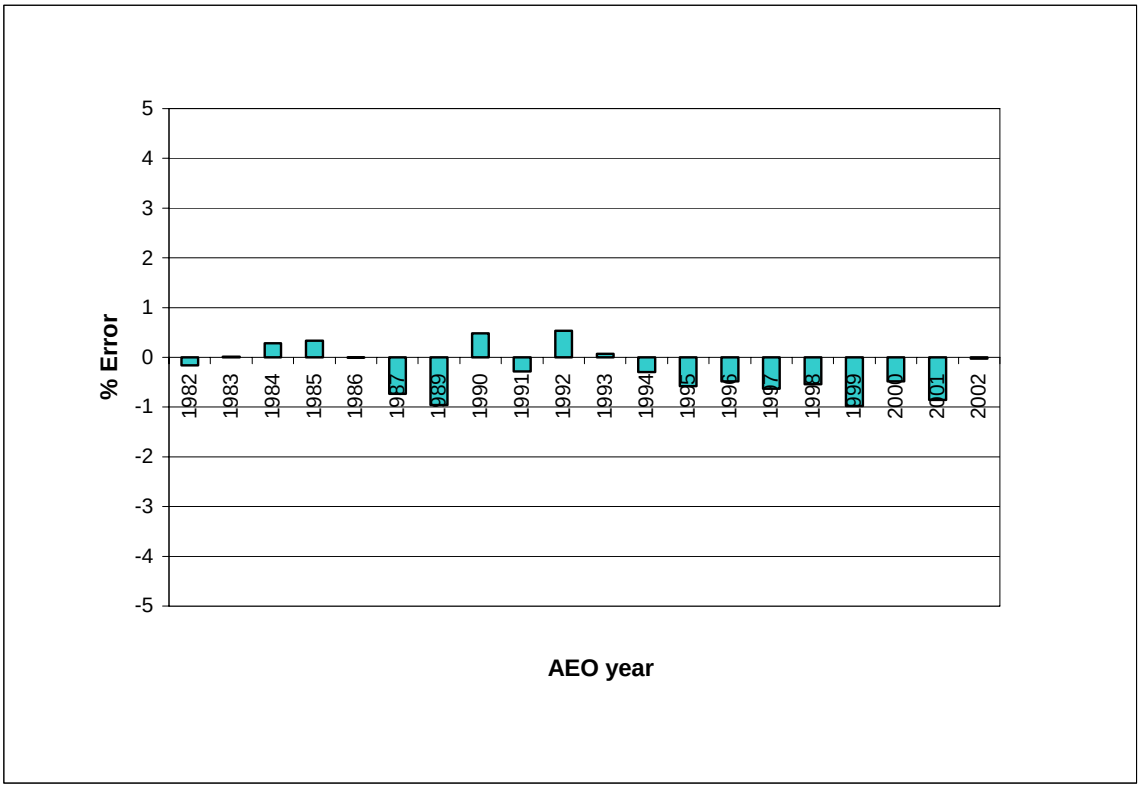

Figure 6. Percentage baseline error in projected energy consumption.

In addition to investigating the sources of error averaged across all projections, it is also worth examining whether later projections are more accurate than earlier ones, due to improvements in data or methodology, or as a result of the change in models from IFFS to NEMS in 1993. To address this question, we examine the accuracy of projections with equal time horizons made in different projection years, using the trend error as the best measure of model performance. For example, Figures 5a, 5b, and 5c show mean absolute trend errors ($MAPE_T$) in projections with three-, five-, and seven-year time horizons, respectively. In general, there is no strong evidence for improvement in projections over time. For three-year projections, trend error remains below 3% for almost all AEO years with no clear pattern of change over time. Similarly, for five-year and seven-year projections, there is no unambiguous pattern of changes in trend error over time. In addition, the accuracy of the projections does not appear substantially different since 1994, when the NEMS model was adopted, as compared to earlier years when the IFFS model was used.

It is possible that projections could be improving not through improvements in the accuracy of the projection modeling, but through reduction of baseline errors. However Figure 6, which shows the baseline error for each AEO year, demonstrates that there is no clear pattern to baseline errors over time. In any case, the baseline errors have been fairly low (<1%), and as noted above have been only a minor contributor to projection inaccuracy.

Thus our analysis of visible errors in consumption indicates that the EIA projections have been accurate to within about 2% over time horizons of less than ten years, with errors climbing to about 4% at time horizons of 10-13 years. Errors are due mainly to inaccurate projections of the trend in consumption (with an important contribution from variability error at short time horizons). Consumption projections have also tended to be too low, on average, and the source of this downward bias is downward bias in projections of the trend. We find no evidence of improvement in projections over time.

Gross Domestic Product and Energy Intensity

We next examine components of visible error in projected energy consumption in a different way, by decomposing it into contributions from errors in forecasting real GDP[3] and energy intensity (EI). Energy intensity is defined as the ratio of energy consumption to economic production, and thus total consumption is just the product of GDP and energy intensity. Figure 7a shows that errors in forecasting GDP and energy intensity contribute about equally to, and are substantially greater than, the errors in energy consumption forecasts. $MAPE_V$ for GDP and EI are about 3-7% up to a nine-year time horizon (compared to 0-2% for errors in energy consumption), and grow rapidly to 10-20% beyond a ten-year time horizon (compared to 3-8% for errors in energy consumption). This suggests that substantial canceling of errors in GDP and EI occurs, so that the smaller errors in consumption are the result of larger, but offsetting, errors in GDP and EI.

[3] An identical analysis based on nominal GDP – the metric used by EIA in its own evaluations (Holte 2001) – does not alter the conclusions we reach, and produces only slightly different quantitative results (slightly larger errors).

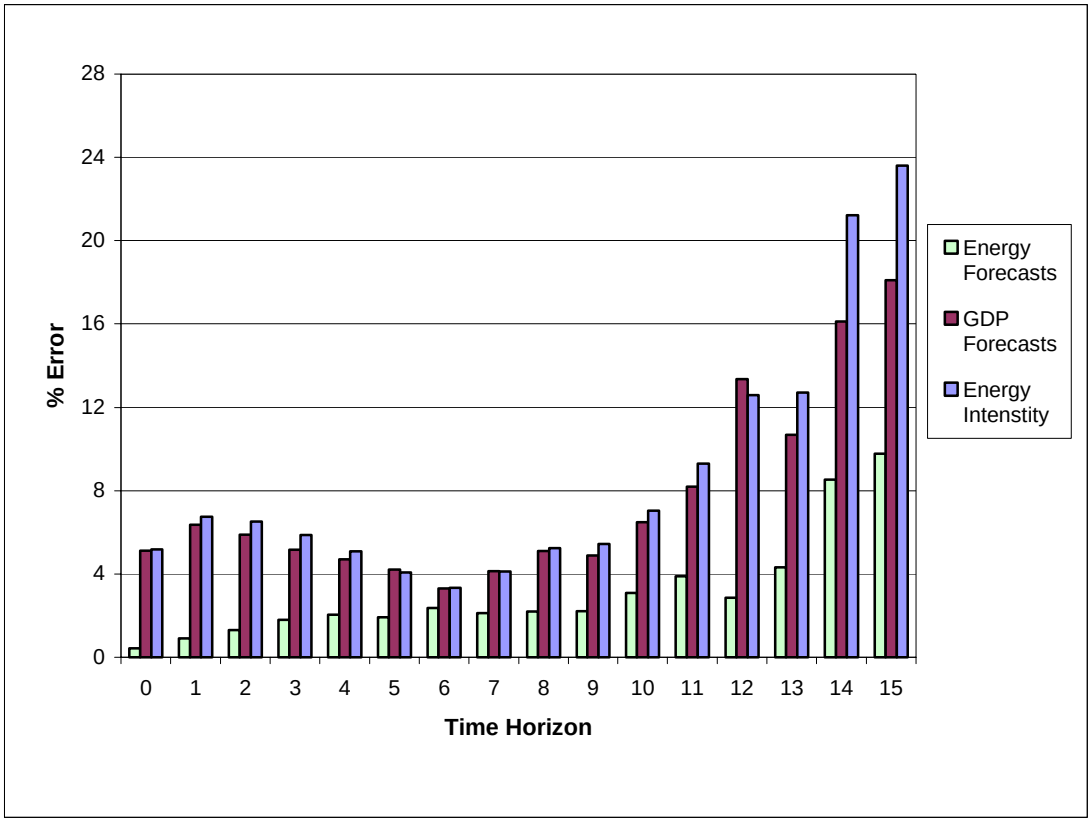

Figure 7a. Mean absolute percentage visible error.

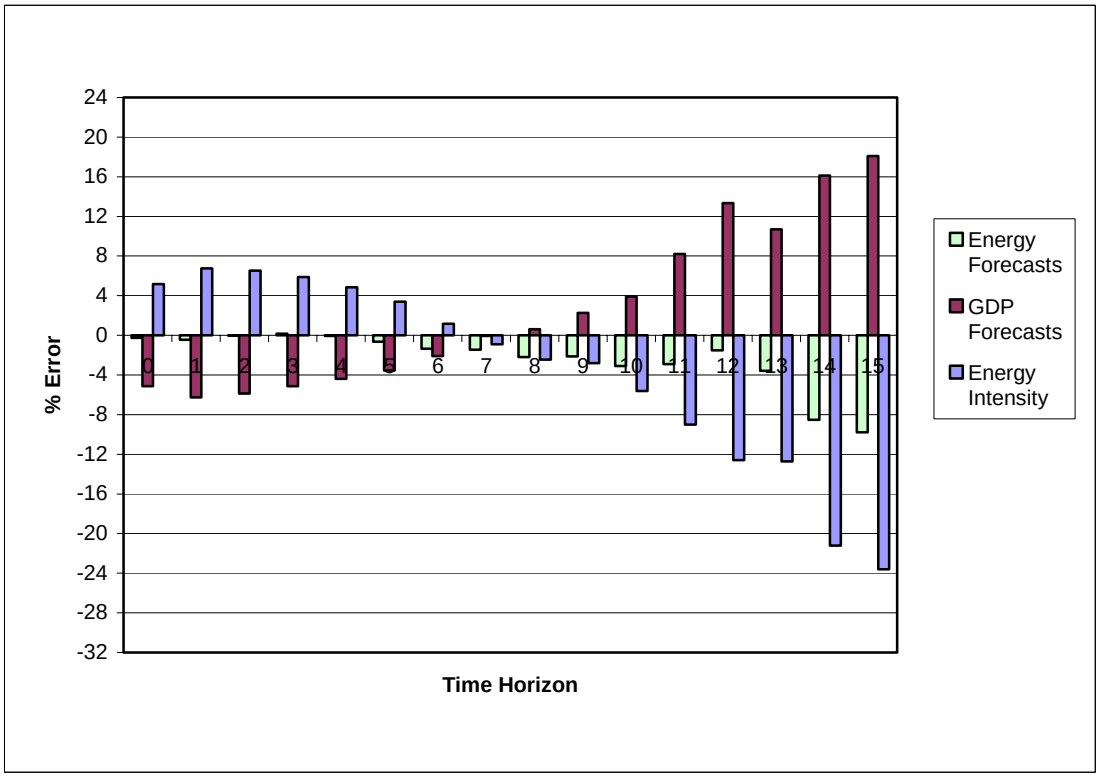

Figure 7b. Mean percentage visible error.

Figure 7b confirms this suggestion. It shows that, on average, errors in GDP and in EI have been of opposite sign. GDP has generally been under-predicted in the short term (time horizon less than seven years), and over-predicted in the longer term; the opposite is true of energy intensity. This general pattern can be explained by two factors, demonstrated in Figures 8 and 9. First, GDP projections are too low, and EI projections are too high, in the short term due to baseline errors. Baseline error in GDP has been substantial (Figure 8a) and consistently negative (Figure 8b) over the entire period, while baseline error in EI has been substantial (Figure 9a) and consistently positive (Figure 9b). Note that the reason the baseline errors in GDP and EI mirror each other is that baseline estimates of EI are derived quantities calculated as the ratio of the estimates of GDP and energy consumption. Because the baseline GDP errors are large and negative, and the baseline consumption errors are small, the baseline EI errors are large and positive.

Second, the reason that projections of GDP are too high, and energy intensity projections are too low, in the longer term can be traced to trend error, and is therefore due to the performance of the projection models themselves. Trend error in GDP grows substantially at longer time horizons (Figure 8a) and is consistently positive (Figure 8b), while trend error in EI grows substantially (Figure 9a) and is consistently negative (Figure 9b).

Both the consistent bias in these projections and their magnitudes have important implications for evaluating the EIA projections. Consider first the over-optimistic projections of the trend in GDP. As discussed earlier, the GDP projections are based on an exogenous forecast supplied to EIA by an outside consulting firm (DRI), which is then modified by feedbacks with the energy model. The over-optimism of the final GDP projections could be due to consistently biased DRI forecasts, or to overly strong modification by energy system feedbacks acting in the direction of increased economic growth. Since the original exogenous DRI forecasts are not available, this question cannot be answered definitively. However, the magnitude of the GDP errors suggests that the feedbacks, which likely constitute a smaller adjustment to GDP growth, are unlikely to be the major part of the explanation. It may very well be that the exogenous GDP forecasts have been the primary source of bias.

Next, consider the consistent under-projection of the trend in energy intensity. At first glance, the rough symmetry in trend errors between GDP and EI might suggest that they are related to each other, much as the roughly symmetric baseline errors in these two quantities are. For baseline error, the EI errors are a simple mathematical consequence of consistently positive errors in estimations of GDP, combined with relatively accurate, and independent, estimates of total consumption. As a result, baseline errors in EI simply mirror errors in GDP, but with opposite sign. However, the situation with trend errors is distinctly different. The projected trend in energy consumption is *not* independent of the GDP projection. In fact, GDP is perhaps the most important driver of consumption. Thus, the trend errors in EI reflect a true shortcoming of the energy projection, and are not a simple mathematical consequence of the trend errors in GDP.

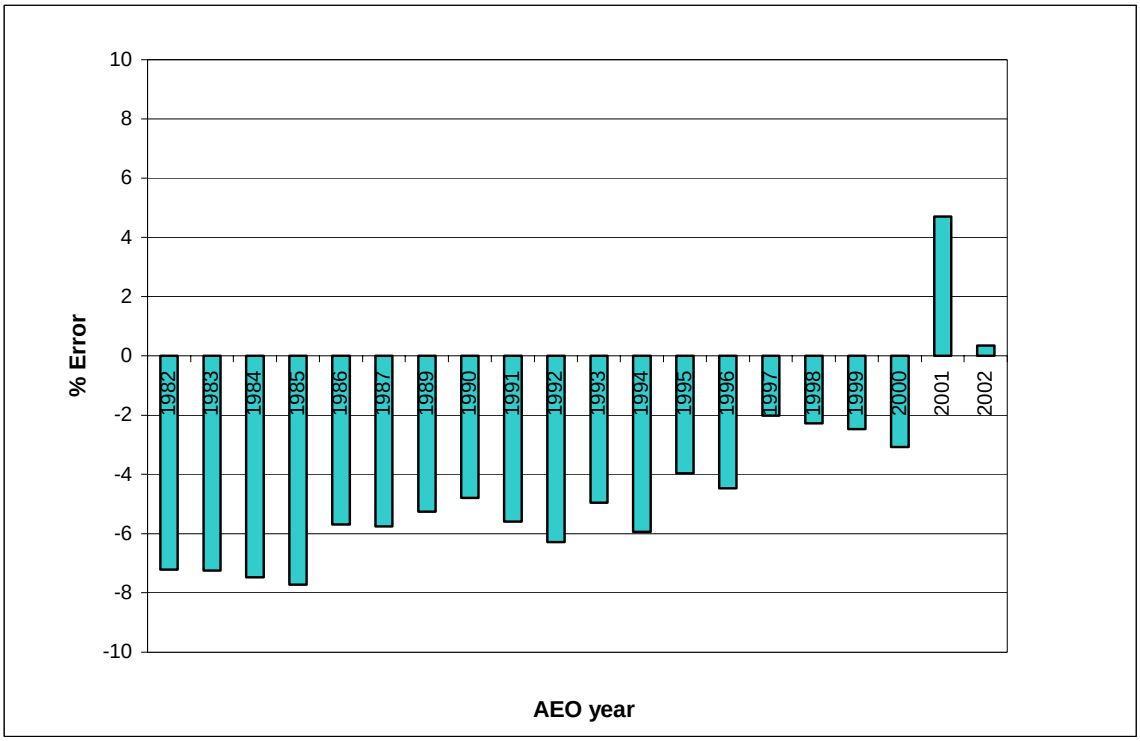

Figure 8a. Percentage baseline error in projected GDP.

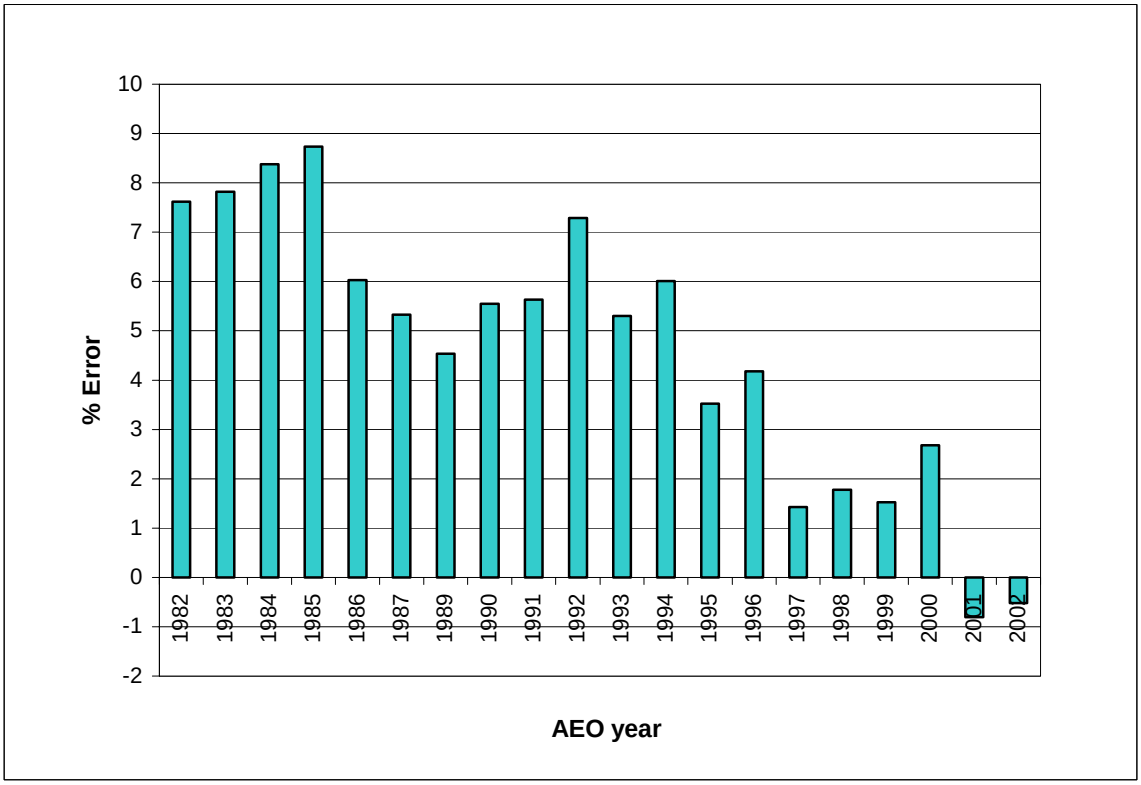

Figure 8b. Percentage baseline error in projected energy intensity.

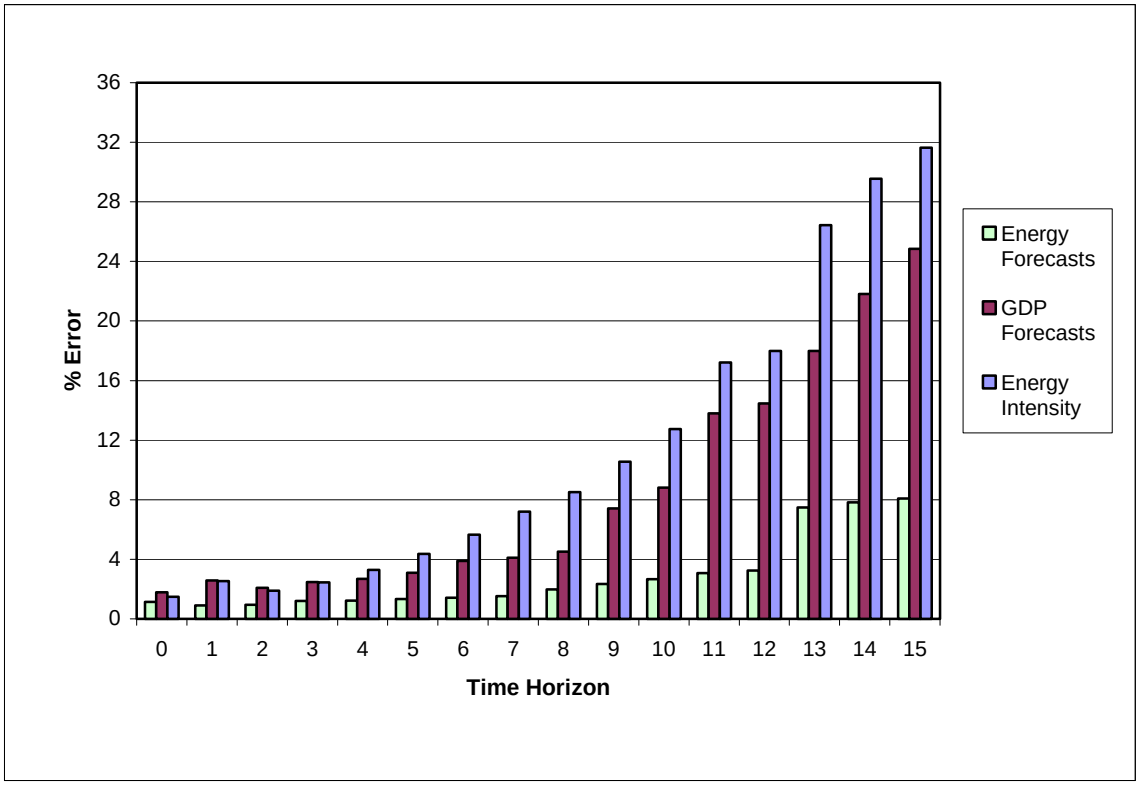

Figure 9a. Mean absolute percentage trend error.

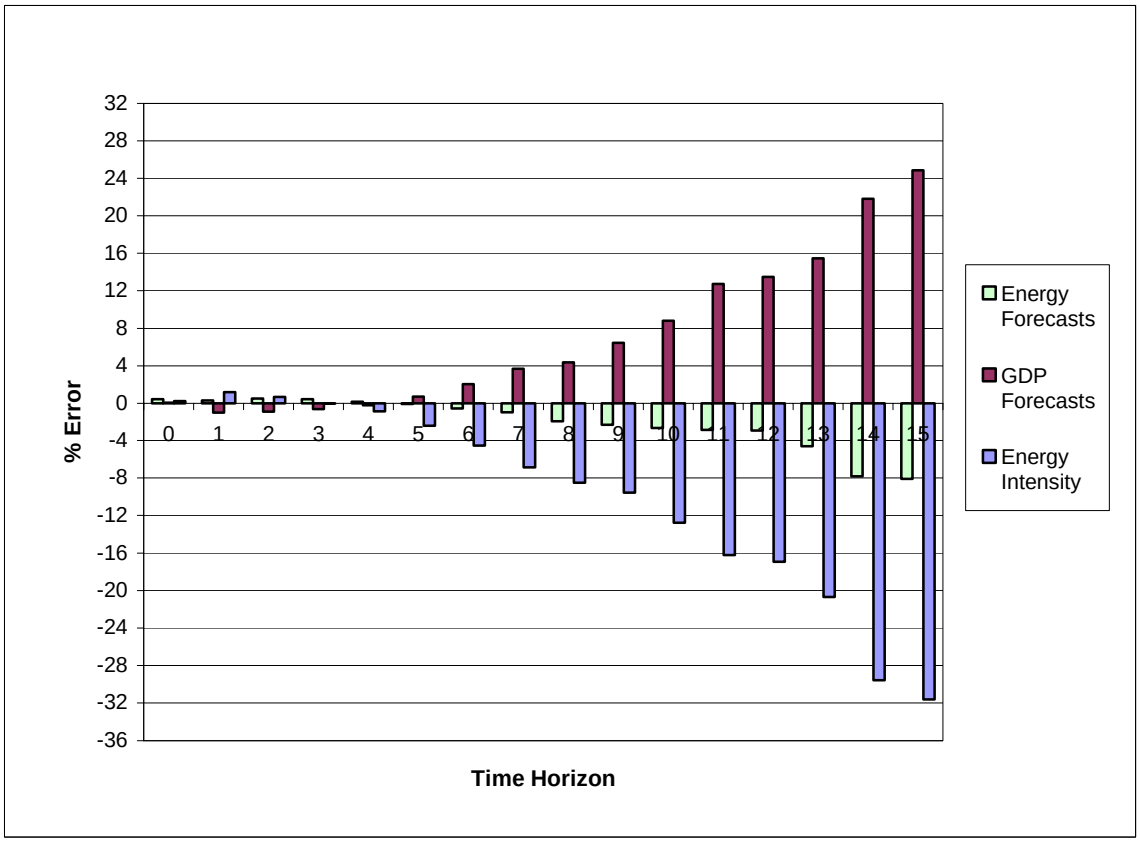

Figure 9b. Mean percentage trend error.

There are at least three possible explanations for the systematic under-projection of EI: the effects of income, prices, or efficiency. It may be that the income elasticity of demand in the projection model has been too low, so that rising incomes in the model do not generate a sufficiently strong increase in energy consumption. It may be that the price elasticity of demand has been inaccurate, so that responses to projected prices have been in error. Similarly, it could be that the projections of the energy prices themselves have been too high (as found by Holte 2001), suppressing the increase in energy consumption in the model below actual consumption levels. Finally, it is possible that projections of the demand for energy services has been accurate, but that projections of the efficiency with which those services are delivered have been overoptimistic, leading to an under-projection of energy consumption. It is beyond the scope of our analysis to pursue these hypotheses here; they remain important questions for future work.

Turning from the bias in the projections to a consideration of the magnitude of the errors, note that the trend errors become quite large with increasing time horizon. For GDP, the error is less than 5% up to a time horizon of about eight years, but grows to exceed 15% beyond a time horizon of 12 years. Similarly, trend errors for energy intensity remain below 5% up to a time horizon of five years, but exceed 15% beyond a time horizon of 10 years. (As shown in Figure 9a, errors in both GDP and EI are even larger at the 15-year time horizon, but the sample size is too small to meaningfully characterize typical performance over these time spans.) If these findings are robust indicators of projection performance, they imply that very substantial inaccuracies are plausible in current projections over time horizons of a decade or more. As previously noted, projections of energy intensity itself play an important role in policy, heightened recently by the Bush administration's climate policy based on goals for improvement in carbon intensity. Equally important is the fact that these results imply that improvements in either GDP forecasts, or in energy intensity forecasts, will lead to potentially large *increases* in errors in projected energy consumption. Errors in consumption forecasts are currently small because they mask large offsetting errors in GDP and EI forecasts. Thus improving only one component will lead to less accurate consumption forecasts.

Finally, we also looked for evidence of improvement in projections of GDP and EI over time. Analysis of $MAPE_T$ for GDP, and separately for EI, for time horizons of three, five, and seven years (not shown) as a function of the projection year show no unambiguous trend toward improvement in forecasts over time. Similarly, there is no indication that GDP or EI forecasts made with NEMS are any more accurate than those made with IFFS (results available from authors).

Conclusions

Our analysis shows that visible errors in projections of U.S. energy consumption have, on average, been too low, but their magnitude has been relatively small (a few percent) up to about 10 years in the future. On the one hand, this level of accuracy stands in marked contrast to the typical level of accuracy in forecasts of energy prices and of macroeconomic growth, which generally fare much worse. On the other hand, we find that the small errors in EIA consumption forecasts are due in part to large offsetting

errors in GDP and energy intensity, which grow to more than 15% at a time horizon of 10-12 years or more.

These errors can inform estimates of the uncertainty in current projections of future energy use. One might be tempted to assume that current projections of consumption ten years into the future may have an uncertainty of a few percent, based on the average of past performance at this time horizon. While we believe our analysis provides a benchmark against which to gauge uncertainty estimates, it must be kept in mind that the future may be harder or easier to project than the past. Given the substantial variations in the magnitude and direction of U.S. energy consumption forecasting errors since the 1960s (Laitner et al. 2003), a change in forecast performance is a clear possibility. Two ways in which the past two decades would no longer serve as a useful guide are (1) if the energy system experiences disruptions, such as those that occurred during the 1970s, but that are not represented in the period of our analysis; or (2) if the large offsetting errors in GDP and energy intensity forecasts no longer tend to cancel each other, as they have over the past 20 years. This caution echoes Landsberg (1985), who reviewed an energy projection he was involved in making in 1963 and found that accurate projections for particular variables were almost always the result of large offsetting errors in the components. “Divining the future correctly in the aggregate can be quite an ego trip,” he wrote, “but its usefulness depends largely on the question one seeks to answer. Nor can you bank on offsetting errors. Errors can also be compounding.” (p.14)

We find no clear evidence of improvements in projections over time since 1982, and no clear difference in projections made with the IFFS model, or its successor, the NEMS model. This is not inconsistent with the conclusion of Cohen et al. (1995), who find evidence for improvement in energy projections made between the late 1970s and early 1980s, given the different time periods of the analyses. Our analysis also suggests some priorities for improvements. First, focusing both on better projections of energy intensity and better projections of GDP would be helpful. GDP forecasts suffer from substantial baseline errors, suggesting that improving the quality of the baseline data in GDP forecasts could contribute to better energy forecasts. It would also be important to take into account “period effects”; i.e., particular times during which prevailing economic conditions may make it easier or harder to forecast GDP (McNees 1992). We also note that variability can be an important source of error for shorter-term projections. This should be recognized in evaluating projection accuracy in any one year: some error should be expected simply because the models are not designed to simulate inter-annual variability. In addition, it may be worth accounting for this fact in setting initial conditions for projections. Currently, the projection model is calibrated to the best estimate of the actual consumption level in the base year. However, it may be advisable to calibrate to the estimated value of the trend in the base year, which will generally be different than the level of actual consumption.

Our analysis also has limitations and could be extended in several ways. We examine consumption, output, and intensity figures aggregated across the economy, rather than by sector. Analysis of the errors in the projections of consumption in individual sectors would give valuable insight in error sources, a strategy that was pursued by Cohen et al. (1995) and Linderoth (2002), who found that relatively small consumption errors in forecasts for OECD countries were typically the result of large, offsetting errors for the transportation and industrial sectors. Such sectoral analyses bear

repeating with the now more extensive set of projections available, and (we suggest) with the methodology used here. In addition, investigating the source of the under-projections of the trend in energy intensity would be important. Additional analysis of price forecasts, measures of energy efficiency, and income and price elasticities of demand would give additional insight into this important question.

References

Bulatao, Rodolfo A. 2001. Visible and Invisible Sources of Error in World Population Projections. Paper presented at the Annual Meeting of the Population Association of America, Washington, D.C., March 29-31, 2001.

Bush Administration. 2002. Global Climate Change Policy Book. Available at http://www.whitehouse.gov/news/releases/2002/02/climatechange.html.

Bureau of Economic Analysis. 2003. National Income and Product Accounts Tables. Updated April 25, 2003. Washington, D.C.: U. S. Department of Commerce. http://www.bea.gov/bea/dn/nipaweb/TableViewFixed.asp?SelectedTable=17&FirstYear=2002&LastYear=2003&Freq=Qtr (cited April 25, 2003).

Cohen, Barry, Gerald Peabody, Mark Rodekohr, and Susan Shaw. 1995. A History of Midterm Energy Projections: A Review of the Annual Energy Outlook Projections. Unpublished paper. Washington, D.C.: Energy Information Administration, U.S. Department of Energy.

Craig, P.P., A. Gadgil, and J.G. Koomey. 2002. What can history teach us? A retrospective examination of long-term energy forecasts for the United States. *Annual Review of Energy and the Environment* 27: 83-118.

Energy Information Administration. 2003a. *The National Energy Modeling System: An Overview 2003.* DOE/EIA-0581(2003). Washington, D.C.: U.S. Department of Energy.

Energy Information Administration. 2003b. *Monthly Energy Review August 2003.* Updated August 26, 2003. Washington, D.C.: U.S. Department of Energy. http://tonto.eia.doe.gov/FTPROOT/multifuel/mer/00350308.pdf, p. 25 (cited September 23, 2003).

Energy Information Administration. 2002. *Annual Energy Review 2001*. DOE/EIA-0384(01), pp. 38, 264-265, 353. Washington, D.C.: U.S. Department of Energy.

Energy Information Administration. 1994. *Supplement to the Annual Energy Outlook 1994.* DOE/EIA-0554(94), p. 1-6, 71. Washington, D.C.: U.S. Department of Energy.

Energy Information Administration. 1983-2003. *Annual Energy Outlook 1982*, annually through *Annual Energy Outlook 2002*. DOE/EIA-0383(82-01), Appendix A. Washington, D.C.: U.S. Department of Energy.

Holte, Susan H. 1997. *Issues in Midterm Analysis and Forecasting*: *Annual Energy Outlook Forecast Evaluation.* DOE/EIA-067(97), pp. 83-111. Washington, D.C.: U.S. Department of Energy.

Holte, Susan H. 2001. *Annual Energy Outlook Forecast Evaluation.* Updated October 2, 2002. Washington, D.C.: U.S. Department of Energy. http://www.eia.doe.gov/oiaf/archive/analysispaper01/forecast_eval.html, http://www.eia.doe.gov/oiaf/archive/analysispaper01/forecast_tbl2.html, (cited March 26, 2002; September 4, 2002; November 2002).

Huntington, H.G. 1994. Oil price forecasting in the 1980s: What went wrong?' *The Energy Journal* 15(2): 1-22.

Keilman, Nico. 1999. How accurate are the United Nations world population forecasts? Pages 15-41 in Wolfgang Lutz, James W. Vaupel, and Dennis A. Ahlburg (eds.), *Frontiers of Population Forecasting*. A supplement to *Population and Development Review*, Vol. 24, 1998.

Laitner, J.A., S.J. DeCanio, J.G. Koomey, and A.H. Sanstad. 2003. Room for improvement: Increasing the value of energy modeling for policy analysis. *Utilities Policy* 11: 87-94.

Landsberg, H.H. 1985. Energy in transition: A view from 1960. *The Energy Journal* 6(2): 1-18.

Linderoth, H. 2002. Forecast errors in IEA-countries' energy consumption. *Energy Policy* 30: 53-61.

Lutz, W., W. Sanderson, and S. Scherbov. 2001. The end of world population growth. *Nature* 412: 543-545.

McCalla, A.F. and C.L. Revoredo. 2001. Prospects for Global Food Security: A Critical Appraisal of Past Projections and Predictions. Food, Agriculture, and the Environment Discussion Paper 35. Washington, D.C.: International Food Policy Research Institute. 71 pp.

McNees, S.K. 1992. How large are economic forecast errors? *New England Economic Review*, July/August, pp. 25-33.

National Research Council. 2000. *Beyond Six Billion: Forecasting the World's Population.* Panel on Population Projections. John Bongaarts and Rodolfo A. Bulatao (eds.). Committee on Population, Commission on Behavioral and Social Sciences and Education. Washington, D.C.: National Academy Press.

Schrattenholzer, L. 1998. A brief history of the International Energy Workshop. Pages 177-185 in J. Weyant (ed.), *Energy and Environmental Policy Modeling.* Boston: Kluwer Academic Publishers.

Shlyakhter, A.I., D.M. Kammen, C.L. Broido, and R. Wilson. 1994. Quantifying the credibility of energy projections from trends in past data. *Energy Policy* 2: 119-130.

Silberglitt, R., A. Hove, and P. Shulman. 2003. Analysis of US energy scenarios: Meta-scenarios, pathways, and policy implications. *Technological Forecasting and Social Change* 70: 297-315.

Smil, Vaclav. 2000. Perils of long-range energy forecasting: Reflections on looking far ahead. *Technological Forecasting and Social Change* 65: 251-264.

Weyant, J., Ed. 2001. *The Costs of the Kyoto Protocol: A Multi-Model Evaluation*. Special Issue of *The Energy Journal*. 448 pp.

Appendix: Adjusting for Dispersed Renewables

In 1990, the EIA began to include "dispersed renewables" in their projections (AEOs) and estimates (AERs) of energy consumption (Holte 1997). Earlier projections and estimates are therefore not directly comparable. Dispersed renewables was an informal term used by EIA to categorize renewables not interconnected to the electric power grid. Dispersed renewables consumption includes total end-use renewable consumption (excluding transportation) and renewable consumption by nonutility power producers. For example, this would include wood used for residential heating and rooftop solar panels for water heating (Eugene Reiser, personal communication).

In order to address this inconsistency for the purposes of their own forecast evaluations, EIA adjusts earlier raw AEO projections with current estimates of past consumption of energy from dispersed renewables. Specifically, a correction factor (*CF*) was added to the raw AEO projections (Eugene Reiser, personal communication), where

$$CF_t = Nonutility\ Power\ Producers_t + Residential_t + Commercial_t + Industrial_t$$

Each term on the right hand side represents total renewables consumption in each sector.

We applied this correction factor to the consumption data taken directly from AEO reports that we used to supplement the data summarized in Holte (2001). Estimates for historical renewable energy consumption by sector were taken from Tables 10.2a and 10.2b of AER 2000 (Energy Information Administration 1983-2003). We confirmed the consistency of this approach with the one taken in Holte (2001) by successfully reproducing their corrected consumption projections for pre-1990 AEOs. The only exception is that for AEO 1989, for projected values in calendar years 1996, 1997, 1998, and 1999, the formula for the correction factor provided by EIA does *not* reproduce projected consumption values presented in Table 2 of Holte (2001). Differences were relatively small; we used the values from Holte (2001).