Energy Efficiency
DOI 10.1007/s12053-008-9015-9

The trouble with energy efficiency indexes: la aritmetic non e opinione

Marvin J. Horowitz

Received: 24 October 2007 / Accepted: 21 May 2008

Abstract Rising energy use and energy-related global greenhouse gas emissions are encouraging new national energy efficiency policies and long-term international cooperative agreements. These require on-going studies to monitor their progress; thus, substantial effort must be directed toward developing reliable evaluation methodologies. One proposed methodology that is gaining in popularity involves energy efficiency indexes. The purpose of this paper is to show that there is a fundamental shortcoming in this approach that makes it unsuitable for estimating policy impacts. This is done in two ways: first, by comparing the calculation of a percent change in energy efficiency indexes to a conventional calculation of a percent change in the level of energy use and, second, by using a Monte Carlo experiment to estimate the probability that policy impacts estimated via an energy efficiency index, even one that has been adjusted, will contain a high degree of error.

Keywords Energy efficiency · Energy efficiency policy · Impact evaluation · Energy index · Energy intensity

Introduction

Energy and environmental public policies, a collective term referring to all governmental efforts irrespective of organizational origin or mode, are expanding in scope and purpose throughout the world. Some of this expansion is due to individual national preferences, but increasingly, it is due to long-term international cooperative agreements. To ensure the success of these initiatives in meeting their goals, there is a growing need for policy monitoring and evaluation research. As stated in the recent Bali Action Plan (2007), the goal of such international agreements is to foster:

> Measurable, reportable, and verifiable nationally appropriate mitigation commitments or actions, including quantified emission limitation and reduction objectives, by all developed country Parties, while ensuring the comparability of efforts among them, taking into account differences in their national circumstances.

In the US, where utility service territory and state and local energy efficiency programs have proliferated since the mid-1970s, there is widespread experience with monitoring and evaluation. The details of many of these studies are be found in the official proceedings of national conferences such as those sponsored by the American Council for an Energy Efficient Economy and the International Energy Program Evaluation Conference. A much smaller number of such studies, typically those with greater rigor and methodological import, are found in respected academic journals, especially those focused on building engineering, transportation research, and energy economics. Yet, on the national level, few energy efficiency policy evaluations exist.

M. J. Horowitz (✉)
Demand Research, LLC,
Fairfax, VA, USA
e-mail: mhorowitz@demandresearch.net

Without much in the way of experience with national energy efficiency policy evaluations either in the US or elsewhere, a method that has attracted a good deal of international interest is the energy efficiency index. Energy intensity, energy indicators, and energy indexes of all kinds have been studied and examined for many years. Engineers and energy professionals employ them to measure manufacturing and services productivity; economists employ them for analyzing energy inputs, outputs, and the components of demand; and policymakers employ them to digest, summarize, and communicate trends about the economy. In many contexts, these statistics have important and valid uses. Indeed, their utility in a broad range of applications is unquestionable.

This being said, it is the purpose of this study to show that the energy efficiency index methodology is not appropriate for evaluating the impacts of national energy efficiency policy. By way of explaining why, the findings of a recent national energy efficiency policy evaluation that employed an econometric approach for estimating energy savings are compared to parallel findings that would be yielded from an energy efficiency index that incorporated the same basic statistics. Using both mathematical comparisons and a Monte Carlo experiment for illustration, this paper demonstrates that the energy efficiency index, with or without adjustment, is prone to error. Given its intrinsic shortcoming, it appears that it would be risky to use this methodology as the basis of a worldwide evaluation system for monitoring international climate change agreements.

Terminology: intensity, indicator, index

Before proceeding, a number of definitions and clarifications are in order. In this paper, unless otherwise noted, *energy intensity* refers to any single ratio in which energy is the variable in the numerator. Hence, energy per person, energy per ton, and energy per dollar are all referred to as energy intensity. This differs from the recent terminology developed by the Odyssee project in Europe (Odyssee 2008), which distinguishes between energy ratios in which a physical object is in the denominator and one in which a macroeconomic variable is in the denominator. For example, Odyssee refers to tons of oil equivalent (toe) per ton of cement produced as *unit consumption*, and kilowatt hours per dollar gross domestic product (GDP) as *energy intensity*. The former is considered by Odyssee to be a bottom-up indicator and the latter a top-down indicator. Also, since in this paper electricity data are all that are referenced, energy intensity will be used synonymously with electricity intensity.

Energy intensity is the kind of a statistical indicator that is popular and useful in everyday life. Like many statistical or economic indicators, it summarizes information and makes otherwise incomprehensible numbers easy to appreciate. Equally important, for those wishing to truly understand the phenomena they represent, they are opportune launching points for serious investigation. For example, the trend in residential energy intensity in the 48 contiguous US states, shown in Fig. 1, is often used as a stepping stone for speculating about the impacts of state energy efficiency policies (e.g., Rosenfeld 2005).

Provided it is cautioned that they do not, in and of themselves, carry information about cause and effect, there is nothing misleading about such indicators. California's flat residential per capita electricity use trend relative to the other states could be due to state energy policies but may also be due to higher state energy prices, differences in state climate, and differences in statewide building materials. Likewise, New York's trend could be due to autonomous factors unrelated to state energy policies, such as differential per capita income or differential demographics, or, it could be due to state energy policies. There is no telling from the indicators themselves.

Statistical indicators of all kinds are mainstays of business, government, and everyday life. Frequently but not always, they are made up of a single variable

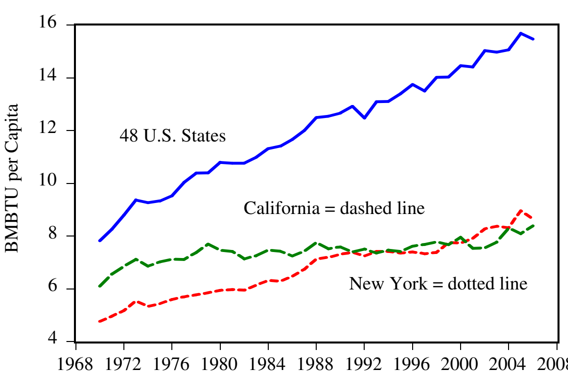


Fig. 1 Residential electricity use per capita, 1970 to 2006

or a single ratio, but sometimes, they are made up of a combination of variables or a combination of ratios. When several variables or ratios are combined into a single value, the aggregate value is often referred to as an *index*. In practice, most index values are normalized to a base period in such a way that the base period value is 100 and the value of any other period is interpretable as percentage changes relative to the base period.

An *energy index* is simply an index composed of variables related to energy. One of the interesting and useful features of all indexes especially energy indexes, is that different arrangements of the terms across periods produce different indexes, each of which is meaningful in its own right. One variety of energy indexes tracks structural changes, another activity changes, and a third, energy intensity or energy efficiency changes. It is the index that tracks energy efficiency changes that is the subject of this paper.

Interest in energy efficiency indexes

Probably no measurement technique has attracted more attention in academic energy economics journals than energy index composition and by extension decomposition. According to an extensive literature review by Ang and Zhang (2000), a total of 51energy index studies were published between the 1970s and 1995, and this number increased by 1.5 times to 124 between 1995 and 2000. Many of these studies analyze and compare the characteristics and virtues of different index formulations, e.g., Laspeyres, Fisher Ideal, or Divisia. Many other index studies are industry-specific, using actual or simulated data to evaluate changes over time in the energy use of, say, pulp and paper processing or steel manufacturing. Moreover, the studies span the globe, employing data from dozens of countries.

Judging by the volume of academic studies and the number of times these studies are referenced, it is no accident that there is growing worldwide interest in energy indexes. According to *Energy Economics* editors Tol and Weyant (2006), starting in 1988, eight out of 10 of their journal's most cited papers are studies of energy index methods and their applications. In addition, papers on this subject have been a staple in the International Association for Energy Economics publication *The Energy Journal*. In 1997, another academic journal, *Energy Policy*, devoted an entire issue to energy and CO_2 emissions index-related papers (Schipper and Haas 1997).

Academic interest in energy indexes stems from the utility of its mathematical character. A common feature of all energy indexes that are aggregated from individual ratios is that they can be disaggregated or decomposed three ways: (a) to track changes in energy use related to changes in the levels of overall activity of the sectors or markets represented in the index, (b) to track changes in energy use that are due to changes in the relative mix or share of each sector or market represented in the index, and (c) to track changes in energy use that are due to changes in input–output relationships. As noted above, the latter is often referred as changes in energy intensity or energy efficiency. Its reciprocal is energy productivity.

Of course, changes in energy efficiency can occur autonomously, that is, due to market and other external forces, as well as due to public policies. Nevertheless, it is this specific property, the isolation of an energy efficiency effect that makes energy indexes attractive for evaluating the impacts of national energy efficiency policies. Witnessing the increasing reliance on indexes for policy evaluation, Ang and Liu (2007) observe:

> Arising from the Kyoto Protocol and the growing concern about world climate change and sustainable development, many countries have been taking steps to reduce greenhouse gas emissions. A common reduction strategy is through taking measures to increase energy efficiency. To evaluate performance, it is necessary to track energy efficiency change and to assess fulfillment of energy efficiency improvement targets on a regular basis and in a rigorous manner. Index decomposition analysis has been used in a number of countries and international organizations to serve this purpose.

One example of such an analysis is Sun (1999) in which index decomposition was used to analyze CO_2 emissions from 1960 to 1995 for the 24 original OECD members. The method employed distinguishes a total international GDP effect, a fuel switching effect, a national GDP share effect, and an energy intensity effect. The study purports to show that, since energy intensity is found to be mostly negative from 1973 to 1993, "This reveals that, after the first oil price shock, policy makers in OECD countries have included improving energy efficiency as part of their economic strategy, and succeeded in that strategy." As is typical in most studies of this kind, changes in energy intensity are not just linked to technical changes in energy productivity but to the impacts of public policies.

Currently, the energy efficiency index approach is positioned to become one of the European Union's main methods for evaluating national energy efficiency and greenhouse gas policies, and for cross-country comparisons of policy compliance and success. Its status is documented in Thomas et al. (2007), a report supporting the European Union Directive on energy end-use efficiency and energy services (European Commission 2006). Also, in a report issued by the European Commission's Intelligent Energy Executive Agency (IEEA 2007), brief outlines are provided of a number of international projects. For one of these, "Evaluation and Monitoring of Energy Efficiency in the New Member Countries and the EU25 (EEE-NMC)," it is stated that, "The project monitors energy efficiency and CO_2 trends, and it evaluates national energy efficiency policy measures." Further:

> It will rely on energy efficiency/CO_2 indicators and a database describing energy efficiency measures by country. The monitoring of energy efficiency progress by sector will combine aggregate bottom-up indicators with detailed indicators to improve the interpretation of the factors behind the trends observed… The data bases will be updated to 2005; new development of methodologies such as the ODEX indicator will be carried out.

According to Bosseboeuf et al. (2005), the aggregated bottom-up energy efficiency index known as ODEX was conceived "in order to meet the political need for monitoring energy efficiency and to have an easily understandable, workable, and comparable indicator depicting the energy efficiency progress in EU member states." As previously mentioned, *bottom-up* refers to unit consumption ratios in which physical variables, such as tons of cement, are in the denominator. In effect, ODEX is an aggregation of microeconomic indicators, as opposed to what Odyssee refers to as a *top-down* index, which is composed of energy ratios with macroeconomic variables in their denominators. Computationally there is no difference between an aggregated bottom-up and a top-down index. In both cases, the indexes are constructed from two or more energy ratios in such a way that structural and activity effects are removed, leaving only the energy efficiency effect.

A similar IEEA project that also involves the development of ODEX, "Monitoring of Energy Efficiency in the EU15 and Norway (Odyssee-MURE)," is described as "Assessment and analysis of energy efficiency improvements and CO_2 abatement (energy related) at the EU and member state level from 1990 to 2005 through updated and harmonized indicators." Enthusiasm for the energy efficiency index method is also expressed by the International Atomic Energy Agency (IAEA) in collaboration with the United Nations, the International Energy Agency (IEA), Eurostat, and the European Environment Agency. It proposes that energy indexes and related analyses be used for promoting sustainable development in emerging economies (IAEA 2005).

In the US, a recommendation of the US National Energy Policy of 2001 was to "…improve the energy intensity of the U.S. economy as measured by the amount of energy required for each dollar of economic productivity" (NEPDG 2001). As a step toward implementing this recommendation, the US Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) gathered more than 150 energy experts from federal, state, and local government and industry, academia, and nongovernmental organizations at its E-Vision 2002 Conference. Its purpose was to discuss ways of reducing US energy intensity. According to the post-conference report by Ortiz and Sollinger (2003):

> For each of four main energy-consuming sectors of the American economy—as well as for the economy as a whole—participants examined historical trends and possible futures in energy intensity; reviewed private- and public-sector experiences at energy intensity reduction; identified current options; and defined goals and the actions necessary to achieve them. This is a key work for those at all levels involved in setting or implementing the National Energy Policy.

Since then, EERE developed a website that came online in 2006. It contains energy intensity ratios for each of the sectors of the US economy, as well as energy indexes.

Some warnings of the danger in using energy indexes for policy evaluation are scattered throughout the literature. For example, Zarnikau (1999) demonstrates how the lack of economic information contained in British thermal unit (BTU) aggregates can distort index decomposition statistics, and Herring (2006) recalls Jevons' paradox to describe how improvements in energy intensity often mask, moreover can even cause, higher absolute energy use. More closely related to the subject at hand, Boyd and Laitner (2001) acknowledge that measured historical trends do not capture policy impacts and illustrate the problem by using estimates of electricity savings attributable to national voluntary public programs to adjust their electricity intensity trend. Golove and Shipper (1997) also acknowledge that energy indexes have shortcomings as measurements of policy impacts. Lastly, Thomas (2005) acknowledges that, in the absence of a hypothetical baseline, energy efficiency indicators cannot distinguish the impacts of energy efficiency policies from stochastic variations in weather, economic growth, and other related factors. Among his recommendations are bottom-up or microeconomic evaluations of individual policies, programs, and services to augment energy efficiency indexes.

Calculating policy impacts

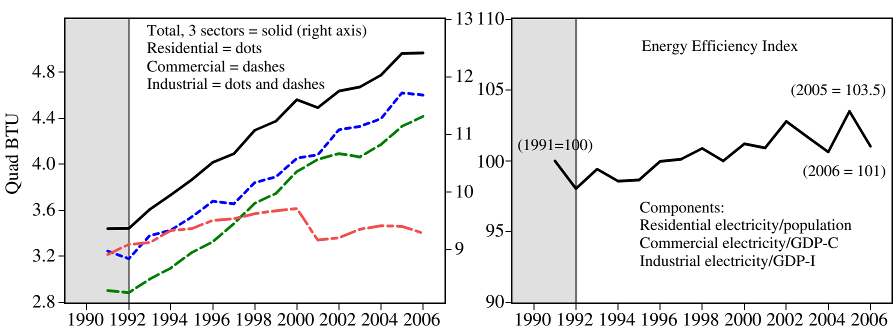
In the academic literature, an energy efficiency index is identified as a component-based, as opposed to an aggregate index, the most well-known among the latter being the simple energy intensity ratio defined as total national energy consumption divided by gross domestic product or E/GDP. Much of the attractiveness of a component-based index approach to policy evaluation stems from its minimal data requirements, simplicity of computation, and ease of analysis, comparison, and communication. Also, a single standardized method that all countries can agree on can eliminate controversy and delay in evaluating energy and environmental policies. In other words, there are many practical reasons for adopting this approach. Unfortunately, none of these reasons matter if the approach cannot do what it is intended to do.

To illustrate how the component-based index approach leads to an energy efficiency index, Fig. 2a shows the electricity consumption trends for the three major sectors of the US economy separately and combined for the 48 US states.

In all, the combined consumption of these three sectors represents more than 95% of the total annual US electricity consumption and, as can be seen, has increased by about 40% from 1991 to 2006. This information can be aggregated, as in Fig. 2b, by an energy efficiency index, revealing that, in some sense, the rise in energy use is not as dramatic as it seems. For this index, the electricity consumption of each of the three sectors is transformed into an energy intensity ratio, and then these ratios are combined into a single index. By subtraction, according to the energy efficiency index, US energy efficiency in 2006 decreased or worsened by only 1% relative to the base year of 1991. Note that in 2005, energy efficiency worsened by 3.5% relative to 1991; to smooth these annual fluctuations, moving averages of the index are often the preferred way to report results.

Because a component-based index combines multiple variables and is normalized to a base period, many different mathematical formulas can be used to compose them, each with unique properties. As noted above, many of the published papers explore and compare these properties. This is not, however, the

Fig. 2 **a** and **b** US electricity use and energy efficiency index (three sectors)



focus of this paper. In this paper, a basic, well-publicized index formula is used for calculating the energy efficiency index (Odyssee 2007). In the three-sector case,

$$\begin{aligned}&\text{Energyefficiencyindex} = \\ &\left[\left(\frac{E_1}{P_1}\Big/\frac{E_0}{P_0}\right)\times\left(\frac{E_1}{E_1+F_1+G_1}\right)\right]+ \\ &\left[\left(\frac{F_1}{Q_1}\Big/\frac{F_0}{Q_0}\right)\times\left(\frac{F_1}{E_1+F_1+G_1}\right)\right]+ \\ &\left[\left(\frac{G_1}{R_1}\Big/\frac{G_0}{R_0}\right)\times\left(\frac{G_1}{E_1+F_1+G_1}\right)\right]\end{aligned} \quad (1)$$

where subtracting 1 from the energy efficiency index allows the index to be interpreted as the percentage change in energy efficiency from the base year. For this study, these terms have the following descriptions:

E	=	commercial sector electricity consumption
P	=	GDP for the commercial sector
F	=	industrial sector electricity consumption
Q	=	GDP for the industrial sector
G	=	residential sector electricity consumption
R	=	US population
Time 0	=	base year 1991
Time 1	=	a single specified year

The specific energy use variables, E, F, and G, were selected because the index constructed from them can be directly compared to the results from a recent econometric evaluation of national energy efficiency policies. The frame variable, P, Q, and R, were selected as denominators merely for their familiarity and frequent use. However, as noted previously, the energy efficiency index formula remains constant regardless of the variables used in the index. This is important to bear in mind because it means that whatever is true about the underlying mathematical character of the index using this example will be just as true for any other energy efficiency index, whether or not the ratios are top-down indicators or unit consumption bottom-up indicators.

Several features of the index calculation are noteworthy. First, the index gives no guidance as to how much of the percentage change in the energy efficiency index is due to autonomous changes versus public policies. Thus, additional information is needed to allocate the changes. However, problems arise even if it is assumed that all of the change is attributable to public policies. This is because, to arrive at the percentage change in each component, the ratio at time 1 is divided by the ratio at time 0, as can be seen in the first term within each bracket. This means that the individual percentage changes are a function of two quantities, one being the percentage change in energy use and the other being the percentage change in the denominator. If it so happens that the percentage changes are the same, then the energy efficiency index will be the same as in the base year even though the policies may have had a large impact. In other words, the second-order change, i.e., the change in percentage change, in energy efficiency will be zero because $\frac{E_1}{P_1}\Big/\frac{E_0}{P_0} = \frac{(E_0\times i)}{(P_0\times i)}\Big/\frac{E_0}{P_0} = X$, where i is a specific percentage change. In both cases, the value of the index will be X.

Unlike the index calculation, in a conventional statistical analysis of policy impacts, the calculation of the percentage change due to the policy involves levels not ratios. Using the notation above, the combined impact of national energy efficiency policies across the three sectors is calculated as:

$$\text{Policy impact}(\%)=\left(\frac{(E_1+F_1+G_1)-\left(E'_1+F'_1+G'_1\right)}{E_1+F_1+G_1}\right) \quad (2)$$

where E'_1, F'_1, and G'_1 represent estimates of the electricity consumption that *would have* occurred in each sector had there not been energy efficiency policies from the base period forward. These three values are called counterfactuals, the business-as-usual scenarios, or simply, the hypothetical baselines. Whether derived from engineering calculations, from metering, from econometric models, from judgment, or whatever, the important point is that the counterfactuals represent the levels of energy use that would have occurred had there been no energy efficiency policies.

Obtaining energy counterfactuals in levels not ratios is the *sine qua non* of energy efficiency policy evaluation. When the policy goal is to reduce energy use, it is energy itself that matters not energy per person, per house, per car, or per dollar. This can be seen most readily by setting the index calculation in time 1, in percent, next to the percentage policy impact calculation, as in the inequality in Eq. 3. One of the noteworthy features of this comparison is not only that the percent policy impact calculation does not contain the denominators P, Q, and R but also that the percent change in the energy efficiency index does not contain the counterfactuals E'_1, F'_1, and G'_1. Rather, the equivalent versions of the counterfactuals are the actual values E_0, F_0, and G_0. In other words, the business-as-usual scenarios in the energy efficiency index calculation are actual energy use at time 0, not what energy use would have been in the absence of national energy efficiency policies at time 1.

$$\left(\frac{(E_1+F_1+G_1)-(E'_1+F'_1+G'_1)}{E_1+F_1+G_1}\right)\neq \left(\begin{array}{l} \quad\left[\left(\frac{E_1}{P_1}\Big/\frac{E_0}{P_0}\right)\times\left(\frac{E_1}{E_1+F_1+G_1}\right)\right] \\ +\left[\left(\frac{F_1}{Q_1}\Big/\frac{F_0}{Q_0}\right)\times\left(\frac{F_1}{E_1+F_1+G_1}\right)\right] \\ +\left[\left(\frac{G_1}{R_1}\Big/\frac{G_0}{R_0}\right)\times\left(\frac{G_1}{E_1+F_1+G_1}\right)\right] \end{array}\right)-1 \tag{3}$$

The importance of this distinction is quite clear when the conventional policy impact calculation using the business-as-usual scenarios is placed side-by-side with the calculation in which the actual base year values are employed as the business-as-usual scenarios.

$$\left(\frac{(E_1+F_1+G_1)-(E'_1+F'_1+G'_1)}{E_1+F_1+G_1}\right)\neq \left(\frac{(E_1+F_1+G_1)-(E_0+F_0+G_0)}{E_1+F_1+G_1}\right). \tag{4}$$

To state the obvious, taking energy use in the base period, time 0, as the counterfactual assumes that the levels of energy use of each component would remain the same if it were not for energy efficiency policy. Unfortunately, absolute levels of energy use at time 1 are more likely to rise, not fall, despite energy efficiency policies. Unless counteracted by even faster growth in the denominators, the end result will be the appearance of a decrease in energy efficiency despite what might indeed be significant policy-related savings.

The observations thus far should be sufficient to demonstrate that changes in an energy efficiency index are virtually unrelated to estimated changes in energy use due to energy efficiency policies. However, this line of thought must be taken a step further to demonstrate that the energy efficiency index approach to policy evaluation has shortcomings even when there is an adjustment to the initial index that, in effect, tries to correct for the inadequacy in the assumed baseline energy use. The most convincing way to do this is to calculate a second energy efficiency index made up of realistic estimates of baseline energy use for time 1. Subtracting the second index from the original one then yields a new, adjusted percentage change in the energy efficiency index. The index-differenced policy impact calculation is:

$$\begin{aligned} &\mathit{Index-differenced\,policy\,impact}\,(\%) = \\ &\left(\begin{array}{l} \quad\left[\left(\frac{E_1}{P_1}\Big/\frac{E_0}{P_0}\right)\times\left(\frac{E_1}{E_1+F_1+G_1}\right)\right] \\ +\left[\left(\frac{F_1}{Q_1}\Big/\frac{F_0}{Q_0}\right)\times\left(\frac{F_1}{E_1+F_1+G_1}\right)\right] \\ +\left[\left(\frac{G_1}{R_1}\Big/\frac{G_0}{R_0}\right)\times\left(\frac{G_1}{E_1+F_1+G_1}\right)\right] \end{array}\right)- \\ &\left(\begin{array}{l} \quad\left[\left(\frac{E'_1}{P_1}\Big/\frac{E_0}{P_0}\right)\times\left(\frac{E'_1}{E'_1+F'_1+G'_1}\right)\right] \\ +\left[\left(\frac{F'_1}{Q_1}\Big/\frac{F_0}{Q_0}\right)\times\left(\frac{F'_1}{E'_1+F'_1+G'_1}\right)\right] \\ \left[+\left(\frac{G'_1}{R_1}\Big/\frac{G_0}{R_0}\right)\times\left(\frac{G'_1}{E'_1+F'_1+G'_1}\right)\right] \end{array}\right) \end{aligned} \tag{5}$$

Of course, it bears pointing out that if E'_1, F'_1, and G'_1 were available, there would be no need to calculate the index-differenced policy impact; the policy impact would already be in hand. However, Odyssee publishes adjusted energy efficiency indexes that are intended to control for national differences in weather, purchasing power, and other nation-specific characteristics. It is thus instructive to see whether this particular adjustment, in which, by definition, E'_1, F'_1, and G'_1 control for nation-specific variables, works as intended. Since the second index should not only make the energy efficiency index impacts more precise but should normalize them so that they can be compared to the similarly derived estimates from other countries, hardly another adjustment could be more advantageous for supporting the use of the energy efficiency index method for calculating policy impacts.

Given this improvement, the fundamental question is whether or not the adjustment is sufficient to correct the shortcomings in the initial index such that the new estimate of the percent change in energy efficiency due to energy efficiency policies is within a reasonable range of the one produced by Eq. 2. There are two ways to attempt to answer this question. One way is theoretical—to try to understand a few of the most important mathematical characteristics of the indexes and thereby gain insight into the conditions needed to approximate the solution to Eq. 2. This exercise is beyond the scope of this paper because the presence of so many variables makes a deductive analysis intractable. Inspection of the calculations makes it readily apparent that, because of the extraneous frame variables, there are unlikely to be many necessary conditions that force the initial or adjusted index to be consistent with conventional policy impact estimates.

An applied or experimental approach is the second way to attempt an answer to this question for this study. This involves a Monte Carlo simulation whose purpose is to measure the size and rate of error between the energy efficiency index impact estimate and the impact estimate derived from a conventional evaluation. This experiment, which yields estimates of the likelihood that the values of the energy efficiency index will fall within reasonable ranges of the true energy efficiency policy impacts, is the subject of the following section.

Measuring the size and rate of index error

For the purposes of the experiment, it is assumed that the values in Eq. 2, the percent change in energy use due to energy efficiency policies obtained from a conventional evaluation, represent *the truth*. This assumption is necessary for comparative purposes only and does not bias the experiment in any way; any reasonable values for Eq. 2 could just as easily serve in the same role. With the values of Eq. 2 as points of orientation, the Monte Carlo experiment is designed to estimate the probability that the energy efficiency index shown in Eq. 1 and the adjusted energy efficiency index shown in Eq. 5 will provide estimates of policy impacts that are within 10%, 25%, or 50% of the true estimates. In other words, the experiment is designed to find out what the level of danger is in relying on an energy efficiency index to estimate policy impacts.

It should be noted that another way to address this issue would be to gather a large number of policy impacts estimated by conventional evaluation methods and compare them to competing policy impact derived from the energy efficiency index method. However, few such impact estimates are readily available. The alternative is this Monte Carlo experiment in which many impact estimates are simulated based on realistic means and variances for all of the components that enter into an energy efficiency index. Such means and variances are available from a three-sector US energy efficiency policy impact evaluation (Horowitz 2007). These impact estimates, which provide estimates of total national electricity savings in 2006 due to energy efficiency policies implemented after 1991, are presently the only US energy efficiency policy impact estimates derived from historical data. For each of the three sectors, national energy efficiency policy impacts are derived from similar models that are estimated using well-accepted econometric techniques.

Taking these impact estimates and designating 1991 as the pre-policy base year, time 0, and 2006 as the treatment year, time 1, the policy impacts for each of the three economic sectors can be described as products of base year random variable multipliers r_E, r_F, and r_G, whose characteristics are displayed in Table 1.

The strategy for the experiment is to compare the values of total cumulative annual US energy savings estimates in 2006—as generated from the sample means and standard errors of the three sectors provided by the impact analysis—to the value that is derived from index construction using the three sectors as components. In other words, using the uncertainty surrounding the counterfactual multipliers for each sector, the findings of many, many versions of the identical impact evaluation are generated. To acquire these new values, r_E, r_F, and r_G were simulated over and over again using the normal distribution.

Table 2 contains the actual data used in this experiment. As before, annual electricity consumption, expressed in BTU, is labeled E, F, and G for the commercial, industrial, and residential sectors, respectively, and the frame variables, which are the denominators used for calculating the indexes, are labeled P, Q, and R for each sector, respectively.

In implementing the experiment, three random variables, A, B, and C, are analyzed. A is the variable whose construction is found in Eq. 2; it represents the total cumulative annual US energy savings estimate in 2006, expressed as a percentage of total electricity consumption in 2006 or time 1 (since there is no

Table 1 Monte Carlo inputs: counterfactual multipliers (1991 to 2006)

Sector	Variable	Mean	Standard error
Commercial	r_E	1.640	0.274
Industrial	r_F	1.469	0.225
Residential	r_G	1.282	0.086

hypothetical baseline in this index, the percentage of total electricity consumption in time 1 is the same as the percentage of total electricity consumption in 1991 or time 0). The construction of B and C are found in Eq. 5, B being the expression on the left side of the minus sign and C being the one on the right side. As previously described, each represents the value of the index from times 0 and 1, and the difference of the two is interpretable as the percentage change of total electricity consumption in time 1. In effect, C is a newly introduced hypothetical baseline that controls for nation-specific characteristics.

Using the terms A, B, and C, it can be seen that if the ratio $A/(B-1)$ or the ratio $A/(B-C)$ is equal to 1, this means that the original index B or index-differenced index $(B-C)$ perfectly match A, the policy impact estimate. In other words, these test ratios are yardsticks of how well the two sets of policy impact estimates match. For example, using the statistics in Tables 1 and 2, it can be seen that A takes a value of -0.099, $(B-1)$ takes a value of 0.010, and $(B–C)$ takes a value of -0.102. This means that the absolute value of the ratio $A/(B-1)$ is equal to 9.533, whereas the absolute value of $A/(B-C)$ is equal to 0.962. Clearly, the test ratio using the index-differenced value very closely matches the true policy impact A, while the original index itself strays from A by more than a factor of nine.

Though this initial result shows close agreement between the index-differenced savings and the true savings, the important issue is not whether the two methods can yield estimates that agree but how often they will agree within certain error bounds. Given the fact that the impact multipliers r_E, r_F, and r_G are measured with uncertainty, A will vary as will C, depending on random elements. Only B will remain constant, since it is exclusively composed of fixed, non-stochastic values.

Procedurally, letting $R=|A/(B-1)-1|$ and $S=|A/(B-C)-1|$, six binomial variables—X_1, X_2, X_3, Y_1, Y_2, and Y_3—are defined, as found in Table 3. In this table, it can be seen that the X's cover a range of conditions for comparing the true percent impact to the index-differenced percent impact, while the Y's cover the same range of conditions for comparison using the initial index. For example, the variable X_1 is constructed from the strictest condition. It takes a value of 1 if S is greater than 10% and 0 if S is less than 10%. Hence, when X_1 is equal to 0.98, this indicates that 98% of the comparisons of A with $(B-C)$ show their values to be within ±10% of each other. Conversely, in only 2% of the comparisons will their values be more than ±10% apart from each other.

The final part of the preparation for the Monte Carlo experiment is to determine the number of trials, n, needed to achieve the means of the X's and the Y's for which there is an error radius of 10% ($\varepsilon=0.1$) at the 95% level of confidence. A trial in this case is defined as a single recalculation of the test ratio from which the means of the X's or the Y's are computed. Determining the number of trials is an essential ingredient in the experiment because too small a number will leave room for doubt as to whether the mean values of the X's and the Y's accurately

Table 2 Monte Carlo experiment inputs: index components

Variable	Description	time 0 (1991)	time 1 (2006)
Consumption			
E	Commercial sector electricity consumption (quadrillion BTU)	2.903	4.762
F	Industrial sector electricity consumption (quadrillion BTU)	3.215	4.112
G	Residential sector electricity consumption (quadrillion BTU)	3.246	4.165
Frame			
P	GDP—commercial sector (trillion, year 2000 $)	4.974	8.864
Q	GDP—industrial sector (trillion, year 2000 $)	1.441	1.587
R	US population (million)	251.274	297.448

Table 3 Binomial variables

Variable	Condition	Value	Condition	Value
X_1	$S>0.10$	1	$S\leq0.10$	0
X_2	$S>0.25$	1	$S\leq0.25$	0
X_3	$S>0.50$	1	$S\leq0.50$	0
Y_1	$R>0.10$	1	$R\leq0.10$	0
Y_2	$R>0.25$	1	$R\leq0.25$	0
Y_3	$R>0.50$	1	$R\leq0.50$	0

represent the value of the actual population means. To calculate the number of trials that would be sufficient to guarantee the validity of the statistical inferences, Bernoulli's theorem (Weisstein 2007) was employed. The theorem shows that for all positive numbers ε, the sample size n will be sufficient to establish a confidence interval for the population mean μ of radius ε and confidence level p when

$$n > \frac{\sigma^2}{(1-p)\varepsilon^2}. \tag{6}$$

A practical difficulty in implementing Bernoulli's theorem is that the standard deviation, σ, needed to determine n is likely to be as mysterious as the population mean μ. Fortunately, because condition (9) is an inequality and not an equation, σ does not need to be known exactly; if σ is estimated with a sufficiently large number, n will be large enough to ensure the validity of the inference.

Based on this property, 50 preliminary experiments, each consisting of a thousand trials, were performed simply to establish a list of 50 standard deviations. From this list, to be conservative, the maximum of the standard deviations was adopted as the underlying standard deviation σ. Furthermore, because random variation might conceivably have produced an unusually small maximum standard deviation in the preliminary experiments, this value was inflated by 10%, and this inflated estimate of the maximum standard deviation σ was used in the place of σ in the condition (9). Based on these procedures, over 57,000 trials were run. This large n more than guaranteed that the means of Y_1, Y_2, and Y_3 would be within an error radius of 10% at the 95% confidence level. Indeed, the actual error radius found for the weakest match criteria, i.e., Y_3, was about 2% at the 95% confidence level.

For calculating the means of X_1, X_2, and X_3, substantially fewer trials were required to achieve this precision level. Indeed, based on each of the 50 preliminary runs, the means of the 1,000 values of X_1, X_2, and X_3 were very close to 1. In other words, R, or the ratio $|A/(B-1)-1|$ was almost never less than 0.5. This preliminary evidence is all that is needed to demonstrate that the initial energy efficiency index B fails as a proxy for A. This comes as no surprise but is a rather remarkable result only because B not $B-C$ is often assumed to be a reasonable estimator of energy efficiency policy impacts. Given these findings, the Monte Carlo experiment needed to go no further for testing the adequacy of B as a measure of true policy impacts.

The estimator $(B-C)$, the index-differenced estimate of policy impacts, is another matter entirely. Recall that using the mean values of the true policy impacts for calculating A and C, the absolute value of $A/(B-C)$ was equal to 0.962, indicating that $(B-C)$ differed from A by only 3.8%. This suggests that $(B-C)$ could be a reliable proxy for A. However, since r_E, r_F, and r_G—and thus A, C, and S—are stochastic variables, a single comparison is far from conclusive in that it does not provide information on how frequently the comparison between the two will be that close.

More trustworthy results are provided by the means of a large sample of comparisons, which is the purpose of the Monte Carlo experiment. As shown in Table 4, the mean values of the variables X_1, X_2, and X_3 do not suggest that the value of the initial comparison is typical. The findings for X_1, the strictest error range, indicates that about 70% of the time, give or take less than 1% at the 95% confidence level, $(B-C)$ is in error by at least 10% relative to A. In other words, the preliminary absolute value of S of 0.038 is misleading. Such close agreement is highly unlikely to occur given repeated sampling of r_E, r_F, and r_G.

It is important to emphasize, however, that a 10% error in the index-differenced policy impact is of no great import. In the US, energy efficiency program impact estimates that are off by only 10% typically

Table 4 Experimental findings for X_1, X_2, and X_3

Variable	Estimated mean
X_1	0.697
X_2	0.396
X_3	0.195

have no effect on the estimated cost-effectiveness of the program or, for that matter, on the perception as to whether or not targeted goals have been met or missed. Therefore, a better perspective on the danger in using an adjusted index to estimate policy impacts is the probability that $(B-C)$ will differ from A by more than 25% or by more than 50%. At these levels, erroneous estimates of policy impacts could have serious implications.

As such, the Monte Carlo experiment finds that about 40% of the time, give or take about 1% at the 95% confidence level, the index-differenced policy impact will differ from the true impact by at least 25%, and that approximately 20% of the time, give or take about 2% at the 95% confidence level, the index-differenced policy impact will differ from the true impact by at least 50%. Put simply, policymakers using estimates derived from an adjusted energy efficiency index, no less a unadjusted index, face substantial, indeed unacceptable, risk of error.

Recommendations and conclusion

It is undeniable that there are many technical difficulties in measuring national energy efficiency policy impacts and that all such efforts bear controversy and uncertainty. Much of the problem stems from the nature of the resource itself. Energy efficiency encouraged through public policies, like any resource savings so encouraged, cannot be empirically observed. What “is” can be counted, what “might have been” cannot, which is why quantifying policy impacts can only be done through inference.

Unfortunately, an energy efficiency index approach to estimating policy impacts does not solve this difficult problem. In fact, national and international adoption of this approach merely for technical and political expedience could be a lose–lose situation in more ways than one. First, because whatever political benefits accrue from using this method to reach speedy consensus will be more than offset by potentially bad impact estimates and potentially bad policy decisions based on these estimates. Second, because bad or misleading estimates and bad policies, once uncovered, will eventually discredit national and international policies of all kinds.

There is a compounded, third loss, too. While minimizing controversy, simplification and standardization of evaluation can stall or even eliminate methodological progress. Like anything that becomes politically entrenched, an incumbent policy evaluation methodology would be exceedingly difficult to unseat, and alternative methodologies would find support and development funding scarce. There are many examples in the US of where this has occurred, a prime one being the adoption two decades ago of California’s standard cost effectiveness tests, which are only now being seriously challenged.

There are no easy answers for how to resolve measurement issues and minimize controversy over the evaluation of policy impacts. However, things are not as bad as they seem. For one, there is much more existing data available for conducting rigorous statistical and econometric national policy evaluations than is commonly believed. The collection of additional primary data—usually the slowest and most expensive data to acquire—is not immediately called for. Organized and coordinated efforts to collect a wide array of secondary data should be the first task of any national or international monitoring effort. To facilitate such secondary data collection, national energy efficiency data centers ought to be created or attached to existing data collection organizations for the purpose of collecting and cataloguing regional and national-level annual data that can be used to analyze and monitor different facets of energy use. In point of fact, much of the data needed for national policy evaluation are already publicly available for many countries from different government, not-for-profit, and private sources. For example, much of the needed data for national evaluations in Europe are currently in the hands of organizations like the IEA, as well as in the databases of projects like Odyssee. It is an oversight to believe that existing data and conventional evaluation methods are insufficient for use in all but the energy efficiency index method.

Second, although there are an infinite number of ways to specify and estimate statistical and engineering models or to conduct qualitative analyses, the range of controversy surrounding most efforts is not as great among experts as it may seem to non-experts. In science, there is always the recognition that there is no one right way to do research; on the other hand, it is equally recognized that there are basic scientific standards that all studies and experiments must conform to, perhaps the most basic of which is making sure that construct validity is not violated.

Last, it seems contradictory to disavow evaluation methods other than the index method as being too complex or too controversial, or simply too impractical, when these very approaches are commonly used for evaluations of individual energy efficiency policies and for analyses of specific national economic issues. If intelligently specified multivariate statistical analyses are acceptable for evaluating a buildings program, the adoption of a new technology, or the effects of tax policy on energy consumption, why should they not be acceptable for evaluating national energy efficiency policies?

On the walls of an Italian pizza place on Ft. Hamilton Parkway in Brooklyn, NY, there is printed, in elegant cursive, an Italian proverb, *la aritmetic non e opinione*, which plainly translates as, "arithmetic is not an opinion." One imagines that this was the owner's advice, perchance warning, to any wise guy or know-it-all patron. Nowadays, it could just as easily serve as advice or a warning to energy and natural resource policymakers worldwide. The trouble with energy efficiency indexes is that they do not and cannot accurately and reliably measure energy efficiency policy impacts. Continued advocacy of this approach as technically and politically expedient is indeed based on expediency, not fact.

Acknowledgment This research was funded by the U.S. Environmental Protection Agency. The author would like to thank Barry Brent for his analytical contributions and helpful suggestions. The author takes sole responsiblity for all errors and opinions found within.

References

Ang, B. W., & Liu, N. A. (2007). Energy decomposition analysis: IEA model versus other methods. *Energy Policy*, *35*, 1426–1432.

Ang, B. W., & Zhang, F. Q. (2000). A survey of index decomposition analysis in energy and environmental studies. *Energy Economics*, *25*, 1149–1176.

Bali Action Plan (2007). The United Nations Climate Change Conference in Bali. Retrieved from http://unfccc.int/meetings/cop_13.

Bosseboeuf, D., Lapillonne, B., & Eichhammer, W. (2005). Measuring energy efficiency progress in the EU: The energy efficiency index ODEX. Proceedings of the 2005 European Council for an Energy Efficient Economy Summer Study, Stockholm, Sweden, pp. 1127–1135.

Boyd, G., & Laitner, J. A. (2001). Recent trends in the U.S. energy intensity: an index number analysis. *International Association for Energy Economics Newsletter*, *10*(2), 4–9 "Skip".

European Commission (2006) Directive 2006/32/EC of the European Parliament and of the Council of 5 April 2006 on energy end use efficiency and energy services (ESD). Official Journal of the European Union 27.4.2006, L 114/64.

Golove, W., & Schipper, L. (1997). Restraining carbon emissions: measuring energy use and efficiency in the USA. *Energy Policy*, *25*(7–9), 803–812.

Herring, H. (2006). Confronting Jevons' paradox: does promoting energy efficiency save energy. *International Association for Energy Economics Newsletter*, *15*(4), 14–15.

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

International Atomic Energy Agency (IAEA) (2005). *Energy indicators for sustainable development: guidelines and methodology*. Vienna, Austria: International Atomic Energy Agency.

Intelligent Energy Executive Agency (IEEA) (2007). *Monitoring and evaluation: 5 innovative projects for an energy-intelligent Europe*. Brussels, Belgium: European Commission.

National Energy Policy Development Group (NEPDG) (2001). *National energy policy: Reliable, affordable, and environmentally sound energy for America's future*. Washington, DC: National Energy Policy Development Group.

Odyssee (2007). Retrieved from http://www.odyssee-indicators.org/Indicators/PDF/odex.pdf.

Ortiz, D., & Sollinger, J. (2003). *Shaping our future by reducing energy intensity in the U.S. economy. Volume 1: Proceedings of the Conference, Prepared for the U. S. Department of Energy*. CaliforniaRAND Science and Technology Policy Institute.

Rosenfeld, A. H. (2005). *Extreme efficiency: Lessons from California*. Sacramento, CA: California Energy Commission.

Schipper, L. & Haas, R. (Eds.) (1997). Cross-country comparisons of indicators of energy use, energy efficiency and CO_2 emissions. *Energy Policy*, *25*(7–9), 639–831.

Sun, J. W. (1999). Decomposition of aggregate CO_2 emissions in the OECD: 1960–95. *The Energy Journal*, *20*(3), 147–156.

Thomas, S. (2005). The quantitative targets for energy savings proposed by the draft directive on energy end-use efficiency and energy services—How to define, how to monitor them. Wuppertal, Germany: Wuppertal Institute for Climate, Environment, and Energy.

Thomas, S., Nilsson, L. J., Eichhammer, W., Vreuls, H., Broc, J.-S., Bosseboeuf, D. et al. (2007). How much energy savings is 1% per year. Evaluation and Monitoring for the EU Directive on Energy End-Use Efficiency and Energy Services (EMEEES). European Expert Workshop, ECEEE Summer Study, La Colle-sur-Loop, FR.

Tol, R. S. J., & Weyant, J. P. (2006). Energy economics most influential papers. *Energy Economics*, *28*, 405–409.

Weisstein, E. W. (2007). Weak law of large numbers. From MathWorld—A Wolfram web resource. Retrieved from http://mathworld.wolfram.com/WeakLawOfLargeNumbers.

Zarnikau, J. (1999). A note: will tomorrow's energy efficiency indices prove useful in economic studies. *The Energy Journal*, *20*(3), 139–146.