nature energy

ANALYSIS
https://doi.org/10.1038/s41560-018-0121-4

Estimation of the year-on-year volatility and the unpredictability of the United States energy system

Evan D. Sherwin[1], Max Henrion[1,2] and Inês M. L. Azevedo[1]*

Long-term projections of energy consumption, supply and prices heavily influence decisions regarding long-lived energy infrastructure. Predicting the evolution of these quantities over multiple years to decades is a difficult task. Here, we estimate year-on-year volatility and unpredictability over multi-decade time frames for many quantities in the US energy system using historical projections. We determine the distribution over time of the most extreme projection errors (unpredictability) from 1985 to 2014, and the largest year-over-year changes (volatility) in the quantities themselves from 1949 to 2014. Our results show that both volatility and unpredictability have increased in the past decade, compared to the three and two decades before it. These findings may be useful for energy decision-makers to consider as they invest in and regulate long-lived energy infrastructure in a deeply uncertain world.

The United States (US) energy system consists of an enormous interconnected network of long-lived infrastructure, which accounts for a large fraction of national greenhouse gas and air pollution emissions, as well as substantial expenditures. Oil and gas extraction alone contributed US$255 billion to gross domestic product (GDP) in 2015, while transportation of goods and passengers contributed US$981 billion[1]. Present investment decisions relating to energy infrastructure will influence the cost, and environmental and health impacts of the US energy system for decades. Understanding how a national energy system is likely to evolve is a difficult task[2], but critically important for informing long-term energy investment decisions. Understanding historical changes in the projected and actual values of key energy quantities can help decision-makers create robust strategies for a deeply uncertain future.

One way to assess the accuracy of past energy forecasts and projections is to perform retrospective analysis. Early work in this field largely began in the 1980s. One approach[3–6] focuses on comparing historical errors from different sets of projections, in this case electricity demand projections from the 1970s, primarily for the purpose of model selection. Another approach[7] describes and attempts to explain historical errors from a set of long-term, national US energy projections using anecdotes.

Further work in the 1990s and 2000s seeks to explain the historical causes of large projection errors. One study[8] attempts to explain large oil price projection errors in the 1980s. Another study[9] computes retrospective errors from the World Input–Output model, created in the 1970s, and discusses reasons for these errors in largely qualitative terms. Every year since 1996, the Energy Information Administration (EIA) has released a retrospective report, detailing its historical projections for 19–21 key quantities from the Annual Energy Outlook (AEO)[10]. These retrospective reports discuss the largest historical projection errors, proposing explanations for these errors, and guidelines for interpreting EIA projections in light of past errors. Other studies[11,12,] attempt to characterize the distribution of projection errors from the AEO.

The early 2000s saw a number of thoughtful review articles comparing numerous sets of past projections[2,13,14]. These articles attempt to convey inherent unpredictability in the energy system and the inadequacy of point projections, with the aim of instilling humility in energy modellers and those who use projections. These articles tend to discuss ways in which projections are useful despite the near-inevitability of large errors. In doing so, they attempt to inform future projection creation processes. Building on these results, one study[15] issues a plea for retrospective analysis of historical energy projections to further inform future projection creation and decision-making practices.

Since then, numerous analyses have sought to assess the historical accuracy of projections, particularly the AEO and World Energy Outlook. The majority of these analyses seek, in some way or another, to determine historical bias, generally on the basis of mean error, or changes in error magnitude over time, using mean absolute error and related metrics[16–23].

The existence of retrospective analyses raises the question of the extent to which insights into past errors can help predict future errors. In short, will the future be as difficult to predict as the past was? We note that the AEO's own low and high oil price scenarios began to widen substantially in AEO 2006, suggesting higher uncertainty in at least that quantity (see Supplementary Note 1 and Supplementary Figs. 1 and 2). AEO projections are ideal for retrospective analysis because they have been produced every year since 1982 by a stable government organization, using consistent methods (see Supplementary Notes 2 and 3), although the process is not stationary in a strict statistical sense (see Supplementary Note 4). The stated goal of the AEO is not to forecast the future, but to project the likely development of the US energy system under the policies in place at the time of the study, and assuming there are no major technological breakthroughs[24]. Still, if the AEO is to guide decision-making, we believe it is important to characterize its historical prediction accuracy.

Here, we attempt to understand whether today's energy system has in fact been harder to predict than in the past by quantifying unpredictability (the frequency of extreme errors) through a retrospective analysis of US energy projections from the AEO reference case. We also investigate whether year-on-year volatility, the frequency of the largest year-on-year changes in key energy quantities, has changed over time. We find that both unpredictability and

[1]Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, USA. [2]Lumina Decision Systems, Inc., Los Gatos, CA, USA. *e-mail: iazevedo@cmu.edu

volatility have been larger in the past decade than in the two and three decades before it.

Year-on-year volatility and unpredictability

We measure changes in the year-on-year volatility and unpredictability for 17 key US energy-related quantities—the price, consumption and production of oil, natural gas and coal; electricity price and sales; residential, commercial, transportation and total energy consumption; GDP and inflation. We measure year-on-year volatility (volatility) by computing year-over-year changes in observed historical values, Δh_t, from 1949 to 2014 for each quantity, computed as:

$$\Delta h_t = \frac{h_t - h_{t-1}}{|h_{t-1}|} \quad (1)$$

where h_t is the historical value of an energy quantity in year t. We identify the years in which the single most positive and most negative changes occur for each quantity as extreme changes. Periods with more extreme changes across different quantities are more volatile. We compare alternative definitions of volatility in Supplementary Note 5 and Supplementary Figs. 3 and 4.

This metric allows comparison of the relative volatility of different quantities, including those with particularly large or small historical variation, for annual-resolution data. Similar metrics are used in the bodies of literature on finance and energy forecasting (see Supplementary Discussion). We define unpredictability as the prevalence of extreme errors within a time period. We define a projection error, $e_{i,t}$, from a projection, $p_{i,t}$, made in year i for year t, as:

$$e_{i,t} = \frac{p_{i,t} - h_t}{|h_t|} \quad (2)$$

where h_t is the historical value in year t.

We define projection length ($l_{i,t}$), with $l_{i,t} = t - i$. As the nearer future may be easier to predict, we perform our analysis over different projection intervals, defined as short-term ($1 \le l_{i,t} \le 5$), medium-term ($6 \le l_{i,t} \le 10$) and long-term ($11 \le l_{i,t} \le 21$). For each projection, $p_{i,t}$, its projection length, $l_{i,t}$, falls into one of these projection intervals, categorizing that projection as short-term, medium-term or long-term. Some grouping is necessary to ensure adequate statistical power in our analysis. In Supplementary Note 6, we define projection intervals in different ways and find that our key results are generally robust to different projection intervals. For each projection interval, we define extreme errors as those errors located outside the 95% probability interval—that is, below the 2.5th percentile or above the 97.5th percentile of errors (see Methods and Fig. 1).

We measure unpredictability by comparing the frequency of extreme errors, f_τ, over a time period, τ:

$$f_\tau = \frac{N_{e,\tau}}{N_{p,\tau}} \quad (3)$$

where $N_{e,\tau}$ is the total number of extreme errors in time period τ, $N_{p,\tau}$ is the total number of projection values in time period τ, and τ is a single year, or a set of years.

Increase in volatility and unpredictability

We compare the relative frequency of extreme changes (year-on-year volatility) and extreme errors (unpredictability) in consecutive ten-year periods.

Our definition of volatility is based on year-on-year changes, shown below for natural gas price and oil production in Fig. 2. Note that the largest decrease in natural gas price and the largest increase

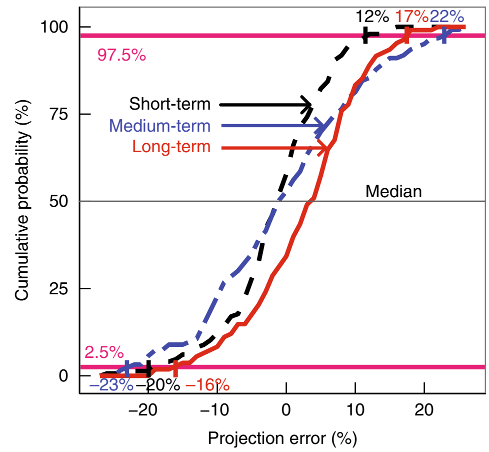

Fig. 1 | Cumulative distribution functions of the projection errors for natural gas production separated by projection interval. The 2.5th and 97.5th percentiles are shown as magenta horizontal lines, whose intersection with the cumulative distribution functions are the extreme error thresholds. Projections 1–5 years into the future are short-term, 6–10 years are medium-term and 11–21 years are long-term. Note that the median value for short-term and medium-term projections is close to zero, while the median for long-term projections is closer to 3%.

in oil production between 1949 and 2014 occur between 2005 and 2014. See Supplementary Fig. 5 for plots of the remaining quantities.

The last decade (from 2005 to 2014) was more volatile than the preceding three decades: comparable levels of volatility are only seen in the 1950s and 1960s. Figure 3 shows extreme changes for each energy quantity over time. The black triangles indicate the year of the greatest increase, Δh_t, in each quantity since 1949, and the red triangles indicate the year of the greatest decrease.

By our definition, there are 34 extreme changes, 2 for each of the 17 quantities. Of these, 9 fall between 2005 and 2014, with only 7 in the entire 30-year period from 1975 to 2004. The remaining 18 fall between 1950 and 1974. Only in the 1950s and 1960s is there a comparable concentration of extreme changes, largely driven by high economic growth rates during that period, a tighter relationship between economic growth and energy consumption, and smaller baseline levels for most quantities.

The few extreme changes that occur between 1975 and 2004 are largely associated with oil and natural gas, due to major swings in international oil markets in the 1970s and 1980s, and rapid changes in the use and regulatory structure of natural gas in the 1980s and 1990s[25].

Of the nine extreme changes in the decade from 2005 to 2014, eight are abrupt decreases, most likely due to the financial crisis and its aftermath. The widespread adoption of horizontal drilling with hydraulic fracturing in shale formations, particularly after 2007, is unquestionably a major factor in the 17% increase in oil production in 2014, and the 54% decline in natural gas prices in 2009.

If we normalize energy production and consumption quantities and GDP, by total US population, the volatility results are similar, with eight extreme changes in 2005–2014 and seven in 1975–2004 (see Supplementary Note 7).

Unpredictability in recent years

We find that unpredictability, measured as the frequency of extreme errors in AEO projections, has increased in the most recent decade. Figure 4 shows the frequency of over-projected (red) and under-projected (black) extreme errors since 1985 (the first year for which there are AEO projections). The placement of circles along the x axis corresponds to the year in which extreme errors occur. The

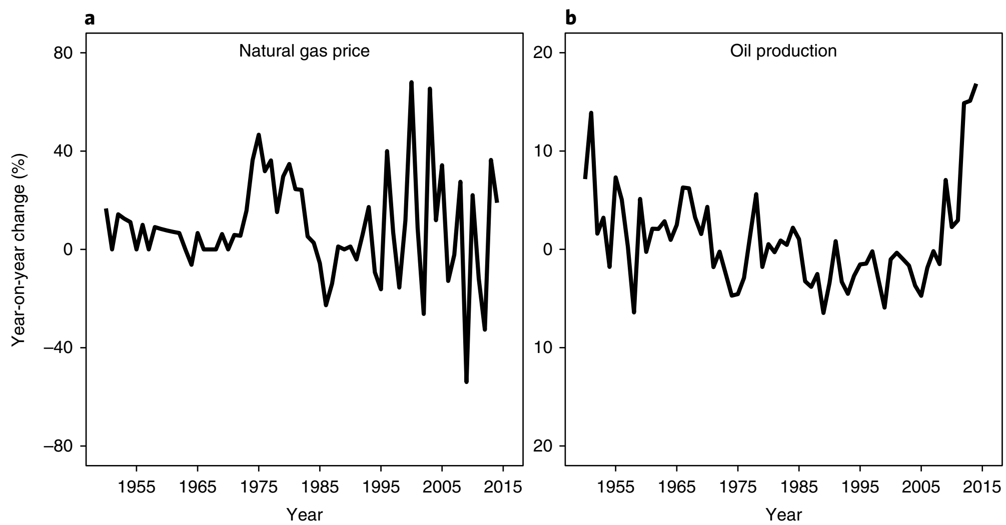

Fig. 2 | Year-on-year changes for two energy quantities. a,b, Natural gas price (in constant 2005 US dollars) (**a**) and oil production data (**b**) show that the largest decrease in natural gas price and the largest increase in oil production both occur between 2005 and 2014.

size of each circle corresponds to the frequency of extreme errors: for example, the 47% frequency of over-projected extreme errors in coal consumption in 2012 means that 47% of all projections of coal consumption for the year 2012 from different AEO reports resulted in over-projected extreme errors. See Supplementary Data 1 for the data underlying Figs. 3 and 4.

The high concentration of extreme errors in the last decade consists largely of under-projections for prices and inflation and over-projections for energy production and consumption. All over-projected extreme errors over the 30-year study period occur in 2005 to 2014 for 10 quantities: production and consumption of natural gas and coal; oil production; electricity sales; total, residential and transportation energy consumption; and GDP. This means that the largest over-projected errors in this period for the short-, medium- and long-term for these quantities were larger than any seen in the preceding 20 years. In the same period, 2005–2014, 15 quantities, all except total and commercial energy consumption, experience an increase in the overall frequency of extreme errors relative to 1995–2004. See Supplementary Figs. 6 and 7 and Supplementary Note 8 for analogues to Figs. 3 and 4 that include three derivative but important quantities, oil and natural gas net imports and energy-related carbon dioxide emissions.

To evaluate the probability of these results occurring by chance, we perform Monte Carlo simulations. In each simulation, we randomly generate 1,000 datasets of projection errors from the 17 quantities. Drawing from a Bernoulli distribution, each simulated projection error has a 2.5% chance of becoming a positive extreme error, and a 2.5% chance of becoming a negative extreme error. The probability of an event occurring by chance, the P value, is the fraction of simulated scenarios in which that event occurred. We replicate each simulation using cross-quantity Spearman correlations derived from the projection errors in each AEO report (see Supplementary Data 2 for these and other related correlations, and Supplementary Note 9 for further discussion of these correlations).

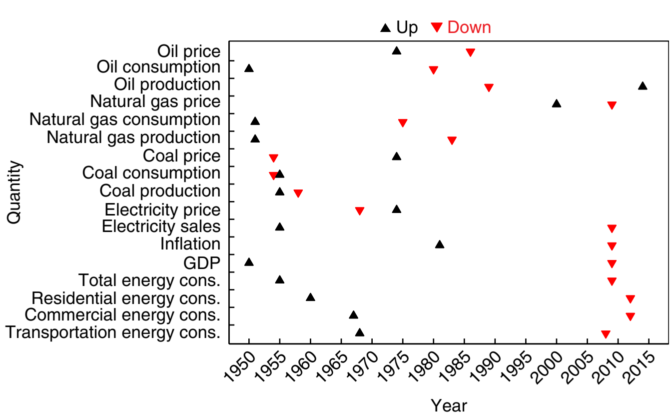

Fig. 3 | Extreme changes for 17 energy quantities, from 1949 to 2014. The black and red triangles indicate the largest year-on-year increases and decreases in the quantities. Note the high concentrations in the 1950s, 1960s and from 2005 to 2014.

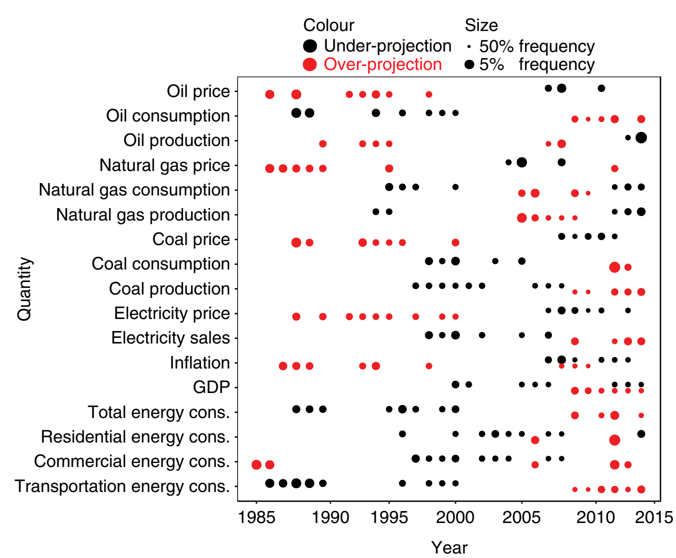

Fig. 4 | Annual frequency of extreme errors for each quantity. The red and black circles correspond to over-projected and under-projected extreme errors, respectively. The size of each circle corresponds to the frequency of extreme errors in that year. Note that for ten quantities, all over-projected extreme errors occur in 2005–2014.

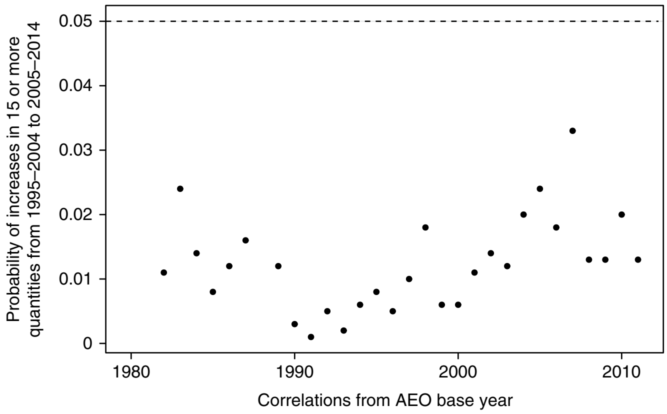

Fig. 5 | The simulated probability of observing increases in the frequency of extreme errors for at least 15 of 17 quantities from 1995–2004 to 2005–2014. This uses correlations from AEO 1982–2011. In all cases, this probability is less than 5% (the dashed horizontal line).

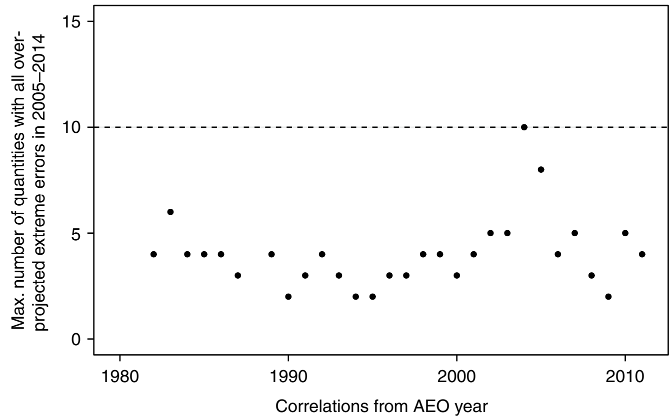

Fig. 6 | The maximum number of quantities for which all over-projected extreme errors occur in 2005–2014. This is derived from 1,000 simulations using empirical cross-quantity correlations from each AEO report. The dashed horizontal line is at the observed value of ten.

We report an upper bound on the highest *P* value across all such simulations as the probability of an event occurring by chance.

Our results suggest that under realistic levels of cross-quantity projection error correlation, it is unlikely but possible, with probability below 5% in all cases, that we would observe an increase in the frequency of extreme errors from 1995–2004 to 2005–2014 for 15 quantities by chance in a time-stationary process (see Fig. 5 and Table 1). It is highly unlikely, with probability below 0.5%, that all over-projected extreme errors would occur in 2005–2014 for ten quantities (see Fig. 6). In other words, both of these results are, in a certain sense, statistically significant. See Methods and Supplementary Note 10 for a further discussion bounding the effects of serial correlation.

Using similar methods, described in Methods, we compute the probability of the volatility results occurring by chance. Given historical levels of correlation between the quantities themselves, Table 2 shows roughly a 10%–20% chance of finding the observed disproportionate clustering of extreme changes in 2005–2014 in a time-stationary process. The large number of downward extreme changes is consistent with extreme errors results, especially the large number of quantities for which all over-projected extreme errors occur in 2005–2014.

Table 1 | Probability of observed results from a frequency of extreme errors analysis under uniform cross-quantity correlation between all unique pairs of quantities

Cross-quantity correlation	*P*(increase in 15 quantities in 2005/2014 versus 1995/2004)	*P*(All + extreme errors for 10 quantities in 2005/2014)
0%	0.1%	0.0%
10%	0.0%	0.0%
50%	1.6%	0.0%
75%	5.7%	0.0%
90%	13.4%	0.0%
99%	30.5%	0.4%

Probability is computed as the fraction of the 1,000 Monte Carlo iterations in a given simulation for which the desired condition holds true. The odds of the first outcome occurring by chance are below 5% for correlations less than 75%, while the second outcome does not occur by chance with probability at least 5% even with 99% correlation.

In Supplementary Note 6, we test the robustness of these results to alternative definitions of extreme error, and subsets of the data. See Supplementary Figs. 8 and 9 for graphical representations of these results.

Discussion

We find an increase in both year-on-year volatility and unpredictability for a broad range of quantities in the most recent decade relative to the immediately preceding decades. In Supplementary Note 11, we demonstrate a case in which considering errors from 2005 to 2014 makes the difference between profit and loss for a liquefied natural gas export terminal (see Supplementary Fig. 10 for the natural gas price scenarios used). Still, volatility was highest in the period from 1950 to 1974, meaning that the relative quiescence from 1975 to 2004 may itself be the anomaly. Also, the observed increase in volatility may be due to increased flexibility in energy infrastructure deployment (for example, distributed energy resources and hydraulic fracturing wells), but the implications for energy decision-makers are largely the same regardless. Note that high concentrations of extreme errors begin before both the massive expansion of hydraulic fracturing and the great recession. For example, a concentration of extreme errors between 2005 and 2006 for natural gas production, consumption and prices is visible in Fig. 4. This suggests that the observed increase in volatility and unpredictability in this decade is due to a number of interlinked, unanticipated developments. While all the authors of this paper have

Table 2 | Probability of observed results from extreme change analysis

Rank correlation	*P*(9 extreme changes in 2005/2014)	*P*(8 downward extreme changes in 2005/2014)
0%	2.1%	0.1%
10%	7.3%	0.9%
50%	18.5%	9.4%
90%	25.6%	15%
99%	27.5%	15.6%
Historical correlations	27.3%	14.9%

Using historical correlations, or cross-correlations above 50%, the probability of each of the two outcomes occurring by chance is well above 5%.

worked on topics related to energy systems, we do not claim to be experts on all the drivers of the observed results. We suggest that these developments may have contributed to the observed results.

In the late 1990s, natural gas was cheap and abundant, and projected to remain so for decades[26]. These expectations, coupled with newly restructured electricity markets in many US states, encouraged construction of natural gas electricity generation plants on an unprecedented scale, particularly between 1999 and 2005[27]. However, large increases in offshore production and Canadian imports of natural gas predicted in the 1990s and early 2000s failed to materialize, driving up natural gas prices[24]. As a result, generation costs for new natural gas plants were higher than anticipated, and plant utilization was much lower[24]. After 2007, tight oil and natural gas production increased massively, driving down natural gas prices, and encouraging increased use of natural gas for electricity generation, displacing generation from coal[24]. Global oil prices rose substantially in the mid-2000s, peaking in 2008, largely as a result of increased demand in Asia, particularly China, and the Middle East[28]. High prices fostered the expansion of unconventional oil extraction in the US and internationally. The financial crisis of 2007–2009 and the ensuing great recession depressed demand for energy[24], placing a downward pressure on energy prices, including oil and natural gas. Vehicle transportation usage declined, with vehicle miles travelled peaking on a per-capita basis in 2005 and on a national total basis in 2007[29]. Finally, industrial energy consumption fell due to deindustrialization and increased energy efficiency, a trend in many sectors of the economy. In combination, these factors led to an unexpected decline in total US energy consumption from its peak in 2007[24].

The observed increase in the volatility and unpredictability of key energy-related quantities may suggest complex structural shifts in the US and world economies and energy systems. Any improvements in the world's most sophisticated energy system models would probably have been overwhelmed by these changes. This turbulence may or may not continue. However, this analysis should serve as a stark reminder of the importance of considering the possibility of further surprises when planning for the future.

Methods

Experimental design. This study aims to identify historical periods characterized by large fluctuations (year-on-year volatility) and extreme errors (unpredictability) for key US energy quantities. We use publicly available historical values and projections for 17 US energy quantities, described below and in Supplementary Methods, to compute these fluctuations and projection errors. We use several non-parametric methods to compare the prevalence of extreme changes and extreme projections errors for these quantities by decade.

Data. All projection data and observed historical values used in the extreme error analysis come from either the Annual Energy Outlook (AEO) retrospective reports, or from the individual AEO reports themselves. The single exceptions to this is GDP, which is derived from a combination of AEO projections of GDP growth and US Bureau of Economic Analysis values of historical US GDP, described in further detail in Supplementary Methods. For a graphical representation of these projected and historical values for all quantities examined, see Supplementary Figs. 11–13.

For our analysis of extreme year-over-year changes, year-on-year volatility, in observed historical values of energy quantities since 1949, we draw data from the EIA's Monthly Energy Review, November 2015 when available[30]. The exceptions are US GDP and inflation, which we draw from the US Bureau of Economic Analysis[31,32]. All prices in the volatility analysis are in nominal US dollars.

Data collection and integration are described further in Supplementary Methods.

Statistical information. We define unpredictability as the frequency of extreme errors for one or several projection years (the projection year would be 2000 for a projection produced in 1990, projecting values for the year 2000).

We define an extreme error relative to its error distribution, the distribution of all projection errors for that quantity. We consider projection error as a percentage, rather than the simple difference between the projected and actual value, because we are interested in the magnitude of the error relative to the observed historical value. We define an extreme error as being outside a specified percentile of the error distribution. In the baseline analysis, we designate as extreme errors all projection errors above the 97.5th percentile, or below the 2.5th percentile. Thus, roughly 5% of all projection values for each quantity are designated as 'extreme errors' (small sample size effects can increase or decrease this rate by up to ±0.6%). We obtain separate extreme error thresholds from the error distributions for each projection interval, short-term (1 to 5 year), medium-term (6 to 10 year), and long-term (11 to 21 year) projections. In Supplementary Note 6, we test the effect of alternative definitions of extreme error on our results.

In Fig. 1 we demonstrate this method of computing extreme error thresholds, using natural gas production as an example. The vertical axis shows the cumulative distribution function of all projection errors for natural gas production since 1985 and the horizontal axis shows the corresponding projection error values. Positive projection errors mean that the projected value was higher than the observed historical value (an over-projection), and vice versa. We show error distributions separately, by projection interval, in black, blue and red lines. The bounds of the 95% probability interval, 2.5th and 97.5th percentiles, are highlighted in horizontal magenta lines. In Fig. 1, we see that in the short term, cases where natural gas production was over-projected by more than 12% or under-projected below −20% are considered extreme errors. In this case, there is little median drift for short-term and medium-term quantities, which have a median at −1.3% error. There is a median drift of 3.1% for long-term projections. For more on mean and median drift, see Supplementary Tables 1 and 2.

See Supplementary Figs. 14–16 for error cumulative distribution functions and extreme error thresholds for all 20 quantities, including 3 derivative quantities: oil and natural gas imports and energy-related CO_2 emissions. See Supplementary Note 12 for a related discussion of mean and median drift among the quantities examined.

We estimate the probability that our main results could have occurred by chance using Monte Carlo simulation, representing extreme errors and extreme changes as drawn from a Bernoulli distribution and an integer uniform distribution respectively.

For extreme errors, we simulate a set of projections for the 17 quantities analysed. We use a Bernoulli distribution to randomly assign each projection as either an over-projected extreme error (probability 0.025), or not (probability 0.975). We similarly assign each projection as either an under-projected extreme error (probability 0.025), or not (probability 0.975). In this way, 2.5% of all projections are under-projected extreme errors, and 2.5% are over-projected extreme errors. For each quantity, we simulate 71 projection errors in 1985–1994, 126 in 1995–2004 and 181 in 2005–2014. These correspond to the number of projections for 16 of 17 quantities. GDP has only 45 in 1985–1994, 124 in 1995–2004, and 181 in 2005–2014. For simplicity, we give GDP the same number of simulated projections as the other quantities. We use Monte Carlo simulation to replicate this process of random extreme error generation 1,000 times per simulation.

We then simulate cross-quantity correlations in two ways. First, we parametrically set a fixed cross-quantity correlation for all pairs of unique quantities for all years. Second, we use cross-quantity correlations derived from projection errors from each individual AEO. We measure cross-quantity error correlation $\rho_{q_i,q_j,k}$, between quantities $q_i \neq q_j$, for a given AEO, k, as:

$$\rho_{q_i,q_j,k} = \mathrm{Corr}_t(\varepsilon_{i,t,k}, \varepsilon_{j,t,k})$$

where $\varepsilon_{i,t,k}$ is the projection error for quantity q, projection year t and AEO base year k. $\mathrm{Corr}_t()$ is correlation, operating over projection years t.

This analysis does not model serial correlation, which, if included, could increase the simulated probability of our results occurring by chance. There are two types of serial correlation of concern. The first is serial correlation between errors from the same AEO for successive projection years. The second is serial correlation between errors for the same projection year from AEO reports from successive years. See Supplementary Note 10 for approximate bounds on the effects of serial correlation.

The two key results we examine are, first, the increase in the frequency of extreme errors in 2005–2014 relative to 1995–2004 for 15 of the 17 quantities and, second, that all extreme errors occur in 2005–2014 for 10 of the 17 quantities. We use Monte Carlo simulation to estimate the probability of each of these events, P(increase in 15 quantities in 2005/2014 versus 1995/2004), and P(All + extreme errors for 10 quantities in 2005/2014), respectively. We estimate these probabilities as the fraction of the 1,000 Monte Carlo iterations in a given simulation in which each respective condition is met.

Figure 5 shows P(increase in 15 quantities in 2005/2014 versus 1995/2004) using cross-quantity correlation derived from projection errors in each AEO report for which there are at least 4 projection errors (AEO 1982–2011). The middle column of Table 1 shows P(increase in 15 quantities in 2005/2014 versus 1995/2004) using parametric cross-quantity correlation. Note in Fig. 5 that for correlations derived from all AEO reports, P(increase in 15 quantities in 2005/2014 versus 1995/2004) is less than 5%.

In the vast majority of simulations, we find that no iteration produces ten or more quantities in which all over-projected extreme errors occur in 2005–2014. This occurs only in 1 of 1,000 iterations using correlations from AEO 2004. Figure 6 shows the maximum number of quantities in which all over-projected extreme errors occur in 2005–2014 for simulations using cross-quantity correlations from each AEO report.

We see instances in which all over-projected extreme errors occur for two to ten quantities, when using empirical correlations from errors in AEO 1982–2011. Thus, the probability of the observed ten quantities occurring by chance is less than 0.5%.

The right column of Table 1 shows the probability of observing ten or more quantities for which all over-projected extreme errors occur in 2005–2014, *P*(All + extreme errors for 10 quantities in 2005/2014), using parametric cross-quantity correlation. The measured probability is zero in all cases except 99% cross-quantity correlation, in which case it is less than 0.5%.

These results suggest that the most extreme errors have indeed become larger for many quantities in the period from 2005 to 2014. We estimate that it is unlikely, but not inconceivable that we could observe our results by chance. We find that accounting for both types of serial correlation described above may increase the probability of an increase in the frequency of extreme errors in 2005–2014 relative to 1995–2004 for 15 of the 17 quantities to above 5%. Adding serial correlation does not increase the probability that all extreme errors occur in 2005–2014 for 10 of the 17 quantities to above 5% unless the values of one or both types of serial correlation are consistently at or above 99%. Spearman serial correlations in both directions described above have a median value of 75%, with a standard deviation of 36%, well below the 99% + level required. See Supplementary Note 10 and Supplementary Table 3 for further details.

Similarly, we simulate the probability that our key volatility results could have occurred by chance using Monte Carlo simulation, assigning an upward and downward extreme change for each quantity using an integer uniform distribution over the years between 1950 and 2014. By definition, there is one upward and one downward extreme change for each quantity over the full study period. Our key results are that in 2005–2014, there are 9 of a total of 34 extreme changes, 8 of which are downward. For each quantity, we randomly select two years from an integer uniform distribution between 1950 and 2014, an upward and a downward extreme change. Note that because the data go to 1949, year-over-year changes begin in 1950. Also, because of sampling with replacement, there is a 1.5% chance of both the upward and downward extreme change occurring in the same year. This will slightly bias our results toward a higher probability of multiple extreme changes in the same decade. We consider cross-quantity correlation between pairs of unique quantities, using both a constant parametric correlation between all quantities, and correlations derived from the historical values of the quantities.

Table 2 shows the probability of both key results occurring by chance, the percentage of iterations in which there are 9 or more extreme changes in 2005–2014, *P*(9 extreme changes in 2005/2014), and the percentage in which there are 8 or more downward extreme changes in 2005/2014, *P*(8 downward extreme changes in 2005/2014). In Table 2, we see that using historical Spearman correlations, both of the baseline results occur with greater than 10% probability, meaning that it is not unlikely that they occurred by chance. The historical correlations are roughly analogous to a uniform correlation level of 50%.

Data availability. Data that support the plots within this paper are available from the US Energy Information Administration (https://www.eia.gov/outlooks/aeo/archive.php), the Bureau of Economic Analysis (https://www.bea.gov/industry/xls/io-annual/GDPbyInd_VA_1947-2016.xlsx) and Oak Ridge National Laboratory (http://cdiac.ornl.gov/ftp/trends/emissions/usa.dat), and are described further in Supplementary Methods.

Received: 30 November 2016; Accepted: 22 February 2018;
Published online: 26 March 2018

References

1. *Value Added by Industry* (US Bureau of Economic Analysis, 2017).
2. Smil, V. Perils of long-range energy forecasting: reflections on looking far ahead. *Technol. Forecast. Soc. Change* **65**, 251–264 (2000).
3. Huss, W. R. Can electric utilities improve their forecast accuracy? The historical perspective. *Public Util. Fortn*. 3–8 (26 December 1985).
4. Huss, W. R. Comparative analysis of company forecasts and advanced time series techniques using annual electric utility energy sales data. *Int. J. Forecast.* **1**, 217–239 (1985).
5. Huss, W. R. What makes a good load forecast. *Public Util. Fortn*. 3–11 (28 November 1985).
6. Nelson, C. R. & Peck, S. C. The NERC fan: A retrospective analysis of the NERC summary forecasts. *J. Bus. Econ. Stat.* **3**, 179–187 (1985).
7. Landsberg, H. Energy in transition: A view from 1960. *Energy J.* **6**, 1–18 (1985).
8. Huntington, H. G. Oil price forecasting in the 1980s: What went wrong? *Energy J.* **15**, 1–22 (1994).
9. Sohn, I. Long-term energy projections: What lessons have we learned? *Energy Policy* **35**, 4574–4584 (2007).
10. *Issues in Midterm Analysis and Forecasting 1996* (Energy Information Administration, 1996).
11. Shlyakhter, A. I., Kammen, D. M., Broido, C. L. & Wilson, R. Quantifying the credibility of energy projections from trends in past data: The US energy sector. *Energy Policy* **22**, 119–130 (1994).
12. Kaack, L. H., Apt, J., Morgan, M. G. & McSharry, P. Empirical prediction intervals improve energy forecasting. *Proc. Natl Acad. Sci. USA* **114**, 8752–8757 (2017).
13. Craig, P. P., Gadgil, A. & Koomey, J. G. What can history teach us? A retrospective examination of long-term energy forecasts for the United States. *Annu. Rev. Energy Environ.* **27**, 83–118 (2002).
14. Bezdek, R. H. & Wendling, R. M. A half century of long-range energy forecasts: errors made, lessons learned, and implications for forecasting. *J. Fusion Energy* **21**, 155–172 (2002).
15. Koomey, J., Craig, P., Gadgil, A. & Lorenzetti, D. Improving long-range energy modeling: A plea for historical retrospectives. *Energy J.* **24**, 75–92 (2003).
16. Linderoth, H. Forecast errors in IEA-countries' energy consumption. *Energy Policy* **30**, 53–61 (2002).
17. Auffhammer, M. The rationality of EIA forecasts under symmetric and asymmetric loss. *Resour. Energy Econ.* **29**, 102–121 (2007).
18. O'Neill, B. C. & Desai, M. Accuracy of past projections of US energy consumption. *Energy Policy* **33**, 979–993 (2005).
19. Winebrake, J. J. & Sakva, D. An evaluation of errors in US energy forecasts: 1982–2003. *Energy Policy* **34**, 3475–3483 (2006).
20. Considine, T. J. & Clemente, F. A. Gas-market forecasts: Betting on bad numbers. *Public Util. Fortn.* 53–59 (July 2007).
21. Fischer, C., Herrnstadt, E. & Morgenstern, R. Understanding errors in EIA projections of energy demand. *Resour. Energy Econ.* **31**, 198–209 (2009).
22. Wara, M., Cullenward, D. & Teitelbaum, R. Peak electricity and the clean power plan. *Electr. J* **28**, 18–27 (2015).
23. Gilbert, A. Q. & Sovacool, B. K. Looking the wrong way: Bias, renewable electricity, and energy modelling in the United States. *Energy* **94**, 533–541 (2016).
24. *Annual Energy Outlook reports, 1982–1987, and 1989–2015* (Energy Information Administration, 1983–2015).
25. *Energy Policy Act Transportation Study: Interim Report on Natural Gas Flows and Rates* (Energy Information Administration, 1995).
26. *Annual Energy Outlook Retrospective Review: Evaluation of 2014 and Prior Reference Case Projections* (Energy Information Administration, 2015).
27. *Form EIA-860 detailed data, 2013* (Energy Information Administration, 2014).
28. Kilian, L. & Hicks, B. Did unexpectedly strong economic growth cause the oil price shock of 2003–2008? *J. Forecast.* **32**, 385–394 (2013).
29. Sundquist, E. & McCahill, C. *For the First Time in a Decade, U.S. Per Capita Highway Travel Ticks up* (State Smart Transportation Initiative, 2015).
30. *Monthly Energy Review* November 2015, 85 (Energy Information Administration, 2015).
31. *Current-Dollar and 'Real' Gross Domestic Product* (US Department of Commerce Bureau of Economic Analysis, 2015).
32. *Gross Domestic Product: Implicit Price Deflator* [A191RI1A225NBEA] (US Bureau of Economic Analysis, 2016).

Acknowledgements

We acknowledge and thank L. H. Kaack of Carnegie Mellon University for sharing the substantial task of data collection and harmonization, as well as J. Apt, M. G. Morgan, A. Davis, W. M. Griffin, E. Rubin, M. Small, G. Wong-Parodi, D. Armanios and S. Feinberg for their thoughtful feedback and advice. The final figures in the Supplementary Information benefited from editorial changes and suggestions from S. J. Davis. This material is based on work supported by the National Science Foundation Graduate Research Fellowship Program under grant no. DGE-1252522. This work was funded in part by the Center for Climate and Energy Decision Making (SES-0949710 and SES-1463492), through a cooperative agreement between the National Science Foundation and Carnegie Mellon University. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Author contributions

E.D.S. and I.M.L.A. secured project funding; E.D.S., I.M.L.A. and M.H. designed the study; E.D.S. analysed the data with iterative feedback from I.M.L.A. and M.H.; E.D.S. created the figures; E.D.S., I.M.L.A. and M.H. drafted and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41560-018-0121-4.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to I.M.L.A.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.