EIA Corrects Errors in Its Drilling Activity Estimates Series

William Trapmann and Phil Shambaugh

Introduction

The Energy Information Administration (EIA) has published monthly and annual estimates of oil and gas drilling activity since 1978. These data are key information for many industry analysts, serving as a leading indicator of trends in the industry and a barometer of general industry status. They are assessed directly for trends, as well as in combination with other measures to assess the productivity and profitability of upstream industry operations. They are a major reference point for policymakers at both the Federal and State level. Users in the private sector include financial analysts assessing investment opportunities. Firms with upstream operations also rely on these data in apprising their circumstances relative those of their competitors.

EIA does not itself collect drilling activity data. Instead, it relies on a commercial source for data on oil, gas, service, and dry well completions, and on well recompletions. These data are provided to EIA monthly on an *as reported* basis. Due to lags in the reporting of well completions which can (though most do not) range up to several years, EIA must statistically adjust the *as reported* completion data to obtain estimates of the numbers of completions that *would have been reported had there been no reporting lags*. Essentially, this is done by assuming that the pattern of reporting lags observed in the past holds true for the present, and making appropriate upward adjustments to the reported numbers of completions on that basis.

As an integral part of its data gathering function, EIA routinely monitors data quality and periodically conducts work intended to enhance its data systems. During a recent effort to enhance EIA's well completion data system,[1] the detection of unusual patterns in the well completion data as received led to an expanded examination of these data. Unfortunately, substantial discrepancies between the data as received by EIA and correct completion counts were identified. For total wells by year, the errors ranged up to more than 2,100 wells, 11 percent of the 1995 total, and the impact of these errors extended backward in time to at least the early 1980s. **Users of the EIA drilling activity data are therefore advised that the drilling activity data which were published or otherwise distributed by EIA prior to February 1998 are substantially in error.**

When the magnitude and extent of the *as reported* well completion data problem was confirmed, EIA suspended its publication and distribution of updated drilling data and EIA staff proceeded to acquire corrected *as reported* files and then revise the statistical portion of its drilling data system to reflect the new, correct information. EIA has now resolved the data problem and generated revised time series estimates for well completions and footage drilled. While the overall industry trends remain consistent with those of the prior, incorrect series, the revised series do exhibit certain differences, chief among which are:

- Drilling activity did attain its peak level in 1981, but the industry completed an estimated 91,469 wells as opposed to the prior estimate of 90,034.
- Gas and oil exploratory wells were greatly under-reported in the post-1985 period with more than half of the wells missing in certain years.
- The decline in drilling in the mid 1990s was not as steep as previously indicated. The drop in prices did lead to fewer wells by 1995, but they had been underestimated by 2,135 wells—a difference of 11 percent.
- Success rates, measured as the share of successful gas and oil wells relative to total wells, improved greatly as early as 1986 as seen in the revised drilling statistics. The prior well data series did not reflect the improvements in exploration until the mid 1990s.
- The relative share of gas and oil wells in successful well counts is comparable.

The remainder of this report first presents background on the drilling activity data: what the records are, how they are collected, and the resulting difficulties in developing timely measures of recent drilling activity. This is

[1]Additional detail on data issues and the Well Completion Estimation Procedure (WELCOM) is provided in a later section of this report.

followed by a discussion of the nature and extent of errors in the raw data files received by EIA. Last, the revised data are presented along with key differences between the prior and revised series and their implications for understanding industry performance.

Drilling Activity Data

About the data. The most widely cited measures of drilling activity consist of summarized information based on individual well records that describe the completion type and status of each well. The individual records contain diverse information regarding the well, including the American Petroleum Institute (API) well number, the well completion date, the well class,[2] the well type[3], location data, and measures of the footage drilled. Drilling data traditionally had been compiled and presented as the records are received, a practice which predates EIA's publication and use of these data. (EIA continued this practice as part of its data operation until the mid 1980s.) This approach, however, reflects the reporting activity as measured by the recipient, rather than the industry's real-time activity level.

Well completion data by report date ("as-reported completions") are not an accurate indicator of actual drilling activity. A preferred measure of drilling activity is a record of well counts and footage drilled by completion date. The completion date marks the point at which the well generally becomes available for production. Drilling measured by completion date is thus more appropriate for industry analysis purposes than drilling measured by reporting date. Well counts by reporting date would match counts by completion date *if* wells were reported with no delay, however, that is not the case. In the early 1980s, EIA staff noticed unusual patterns in the as-reported completion data, which reflected distortions due to a variable and sometimes very extended reporting lag.

The reporting lag often creates incorrect magnitudes for peaks, troughs, or changes in drilling activity. The data as reported also can obscure the timing of these events. The lags are particularly troublesome at times when the drilling trends shift. The historical data provide some especially clear examples of the distortions that can be caused by recording wells by report date. Wells aggregated by completion date now show that total completions peaked in 1981 at 91,469 and had fallen to 84,299 by 1982 (Figure SR1). This pattern reflects the effect of oil price levels on drilling activity and is consistent with the data on rotary rigs running, both of which peaked in 1981. By comparison, total well counts as reported in 1981 and 1982 were 78,538 and 85,795, respectively.[4] The higher count in 1982 reflects an influx of reports of drilling activity that occurred in 1981 and earlier years. The reporting lags were sufficient to cause the 1981 count to be off by roughly 13,000 wells, a 14 percent discrepancy, with the result that (on the basis of the as-reported well completions) the wrong year is indicated as the peak year for drilling activity. Another critical example of reporting lag distortion is the measure of drilling in 1986, when the industry suffered a severe collapse in oil and gas prices. The number of wells actually completed in 1986 were about 40,000, while the number of well completion records received was over 58,000. Reliance on the as-reported well counts would have significantly masked the serious impact of the price decline on the industry. By that time, however, EIA had taken steps to convert its drilling activity series to an as completed basis, which avoided this problem.

Adjusting for the reporting lag. Concerns about drilling activity data arose at EIA in the early 1980s. These concerns resulted in a plan to develop a procedure that would allow aggregation and reporting of the data on a completion date basis. As noted earlier, EIA does not collect well completion data directly, but relies on a vendor to collect the data, which EIA purchases as a monthly compilation. These are the only such data available, so attempts to convert to a completion date basis must utilize these data. Further, the quality of the EIA's drilling activity estimates is directly dependent upon the quality of the received vendor records.

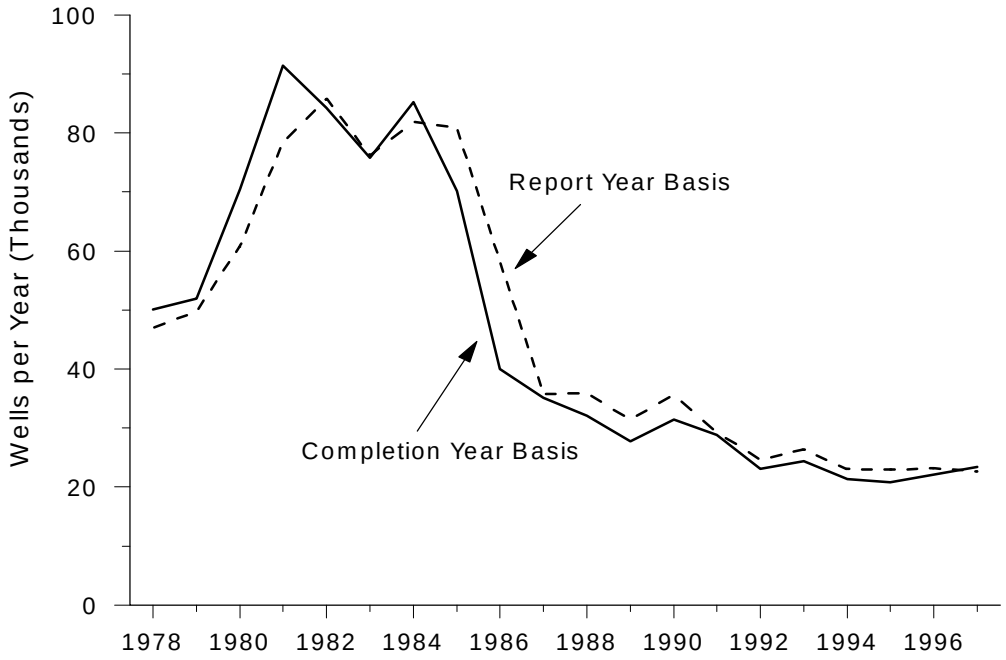
The conversion to completion date required the development of a statistical model with which the as-received completions data could be adjusted for incomplete reporting due to the time lags between completion and reporting. The reporting lags result in cumulative as-reported well counts that are quite incomplete in most months. For example, the data reported in 1986 show that only 14 percent of wells are reported in the same month that they are completed. Twelve months after completion, almost 23 percent of the well completion records had not been received. Even after 60 months of reporting, 2.3 percent of the well completions had not been reported (Figure SR2). These

[2] *Well class* is either exploratory or developmental. Exploratory wells are identified further as either new field wildcat, new pool wildcat, deeper pool test, shallower pool test, or extension well (American Association of Petroleum Geologists (AAPG) well classification codes 1 to 5).

[3] The three *well types* are oil, gas or dry. By convention, wells with both oil and gas zones are categorized as oil completions.

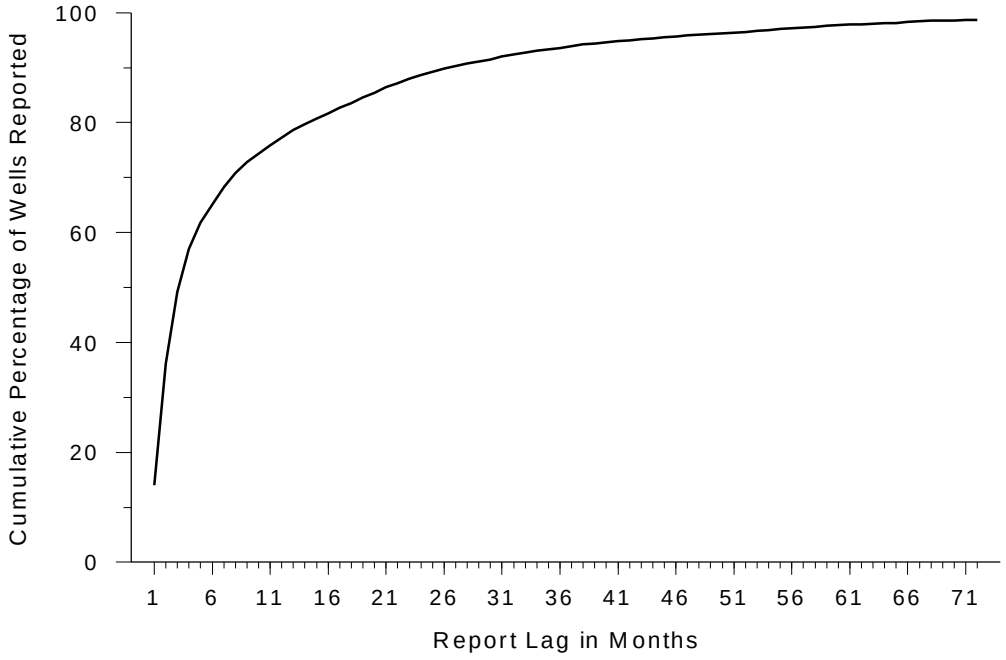
[4] *Monthly Energy Review*, EIA (DOE/EIA-0035), January 1984.

Figure SR1. Well Counts by Completion Date and Report Date



Source: Energy Information Administration, Office of Oil and Gas.

Figure SR2. Reporting Lags for Wells Reported in 1986



Source: Energy Information Administration, Office of Oil and Gas.

reporting lags affect each of the six subcategories (well type by well class), although the effects are variable among the subcategories.

EIA's efforts to convert the well completion statistics from an as-received date to a completion date basis resulted in the Well Completion Estimation Procedure (WELCOM). WELCOM is a system that summarizes the historical records and produces estimates for drilling activity in the most recent years. Estimates are monthly at a national level for six major categories: oil-exploratory, oil-development, gas-exploratory, gas-development, dry-exploratory, and dry-development. Analysis of the data available by the mid 1980s led to the conclusion at that time that the data are fairly complete by 60 months after the actual completion date. Thus, WELCOM uses the cumulative recorded well counts, along with data on rotary rigs running, to construct estimates of the actual numbers of wells completions, and footage drilled, during the most recent 60 months of activity. EIA has used WELCOM since March 1985 to provide the drilling activity estimates published in the *Monthly Energy Review* and the *Annual Energy Review*, as well as other EIA reports. The same drilling data and estimates have also been a part of the foundation underlying numerous analytical efforts including the *Annual Energy Outlook*, the *Short-Term Energy Outlook*, and *Performance Profiles of Major Energy Producers*.

Over the years, EIA's operation of WELCOM has included annual reestimation of the system coefficients as well as minor adjustments to the overall implementation of the methodology. The basic system remains essentially the same as that which was developed in 1984. Since then, the patterns and attributes of drilling have substantially evolved as a result of various regulatory reform initiatives and the shift to generally lower oil and gas prices since the mid 1980s. The collection and processing system for the well records has itself changed. Other changes, such as the fact that the data for active rigs now identifies them by whether they are targeted to oil or gas, have offered opportunities to refine the estimation procedure and enhance the precision of the estimates.

A project to enhance the WELCOM procedure was undertaken by EIA in 1997. Its goal was to take maximum advantage of available data, modify the model specification as appropriate, and test alternate statistical approaches to the estimation of the model's coefficients. The associated data work had an unintended impact, however, when curious patterns in the reported well counts were noticed. Well counts can be and often are highly variable between months. Reported well counts for some months in 1995 and 1996, however, were about half the counts of the prior and succeeding months.

Collaborative examination of the data with the current vendor verified the existence of errors and omissions in the data files provided to EIA in these periods. Review of other time periods disclosed errors that were pervasive in the data over an extensive period. The data files since 1987 were missing some records and contained duplicates of others, updates to many records were not passed along, and records for recompletions -- which weren't expected -- were present in files for some months but not in others.

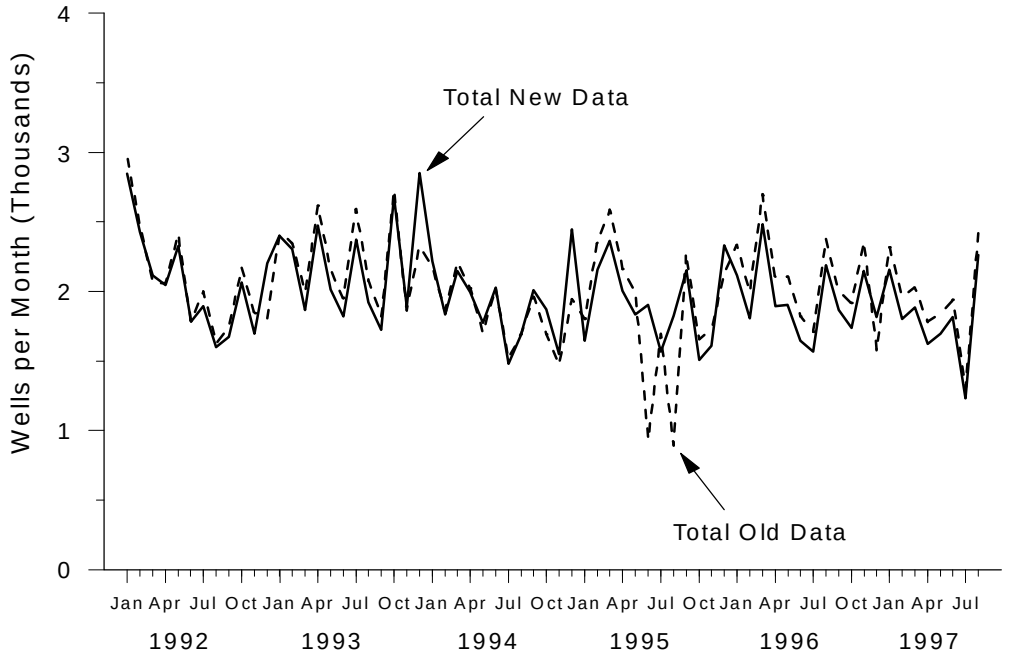
While ready detection of large discrepancies would be virtually certain in a typical EIA data system, the well completion data have a number of attributes that hinder detection. First, the work with the raw data and all initial processing of it are conducted outside EIA. Such an 'arms-length' relationship with the basic data does not allow complete familiarity with the data, and it hampers efforts to investigate concerns with the data. Second, the well completion data are inherently quite erratic, so even sizeable fluctuations are not *prima facie* a cause for concern. A comparison of the well counts as initially received and the corrected counts for 1990 to 1994 show only slight discrepancies, with few exceptions, for the years immediately prior to 1995 (Figure SR3). Yet another factor that may impede ready detection of data problems is the statistically-based processing of the data. The limited numbers of wells reported early after completion are 'inflated' by large factors, which in itself will tend to obscure reflection of data problems in the resulting estimates. Last—and arguably paramount—there is no other timely source of information that can be used to validate the data going into or the estimates coming out of WELCOM.

Resolution of the problem required acquisition of the corrected data from the vendor and re-estimation of the WELCOM coefficients. This has now been accomplished, and EIA has resumed publication of drilling activity data. The revised series exhibit a number of key differences from the prior erroneous series, which may affect perception and understanding of the industry.

Comparative Assessment of the Revised and Prior Estimates

Differences between the prior and revised data series. Comparison of the corrected drilling activity estimate series with the prior erroneous series shows that the overall trends are comparable, with a few significant exceptions. Drilling peaked in 1981 at 91,469 as opposed

Figure SR3. Monthly Well Counts as Reported, 1992-1997



Source: Energy Information Administration, Office of Oil and Gas.

to the previous measure of 90,034 (Table SR1). At least as significant is the drilling pattern of the mid 1990s. The prior estimates for the 1993 to 1996 period showed that the low oil and gas prices through 1995 contributed to annual well completion declines of 17 percent in 1994, then 5 percent. The 1996 surge in oil and gas prices led to a 22 percent rise in drilling activity according to the earlier data. Unfortunately, it appears that estimates based on the flawed data overstated the industry reaction in both directions. The revised drilling statistics show that the relative worsening of economic conditions in 1994 resulted in a decline of 12 percent in wells completed, followed by a 3 percent decline in 1995. This casts a much different light on the industry's responsiveness to the economics of the time. The revised 1995 estimate of 8,252 gas well completions indicates a more moderate reaction to the worsening economic conditions. In total, the revised series shows 2,135 more wells were drilled in 1995 than previously estimated. The lesser decline in 1995 wells also results in a smaller relative recovery in response to the rise in both oil and gas prices in 1996, 7 percent in contrast to the 22 percent change indicated by the prior series.

A number of significant differences are apparent in the exploratory well counts. The key change is that many more oil and gas exploratory wells were completed than previously estimated. This difference appears in the data for the period 1985 through 1995. The shift in both oil and gas completions between series is sizeable, with more than twice as many exploratory wells having been completed in most cases during 1989 through 1992. Dry holes do not shift correspondingly, so the associated success rates[5] are much higher than previously believed. According to the revised drilling statistics, the earlier peak success rate of 30 percent was surpassed by 1986, and success rates have been sustained above 40 percent since 1989. The improvement in industry performance evidenced by this substantial rise in success rates was indicated to be a relatively recent phenomenon according to the previous data series, with rates below 30 percent until 1994. Thus, the corrected drilling activity data provides a fundamentally different picture of industry performance and the investment incentives for firms engaged in exploratory activities (Figure SR4).

Most other aspects of the aggregate well counts, such as the relative numbers of oil and gas wells, are comparable between the prior and corrected drilling activity data series. While most characteristics of the drilling data are consistent between the two series, the changes in drilling levels for certain years and the new measures of exploratory success are so fundamental to an

[5] *Success rates* are measured as the ratio of successful oil and gas completions relative to the sum of oil and gas completions plus dry holes.

Table SR1. Oil Wells, Gas Wells, and Dry Holes, Estimates Before and After Data Correction

Estimates Prior to Data Correction

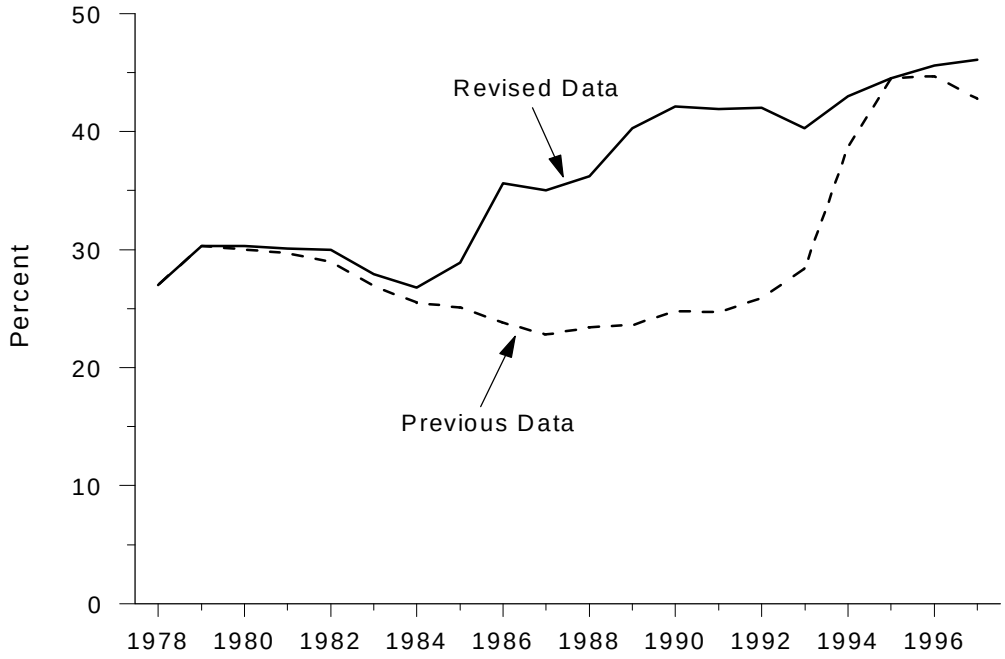
Year	Exploratory				Developmental				All Wells			
	Oil	Gas	Dry	Total	Oil	Gas	Dry	Total	Oil	Gas	Dry	Total
1978	1,191	1,792	8,054	11,037	17,874	12,613	8,537	39,024	19,065	14,405	16,591	50,061
1979	1,335	1,920	7,478	10,733	19,368	13,250	8,560	41,178	20,703	15,170	16,038	51,911
1980	1,781	2,094	9,035	12,910	30,497	15,129	11,302	56,928	32,278	17,223	20,337	69,838
1981	2,667	2,533	12,297	17,497	40,176	17,374	14,987	72,537	42,843	19,907	27,284	90,034
1982	2,470	2,168	11,346	15,984	36,672	16,776	15,036	68,484	39,142	18,944	26,382	84,468
1983	2,113	1,660	10,271	14,044	35,086	12,896	14,065	62,047	37,199	14,556	24,336	76,091
1984	2,335	1,599	11,482	15,416	40,250	15,413	14,315	69,978	42,585	17,012	25,797	85,394
1985	1,879	1,282	9,445	12,606	33,142	12,970	11,763	57,875	35,021	14,252	21,208	70,481
1986	988	733	5,511	7,232	17,713	7,402	7,255	32,370	18,701	8,135	12,766	39,602
1987	859	673	5,179	6,711	15,327	7,084	6,302	28,713	16,186	7,757	11,481	35,424
1988	792	663	4,766	6,221	12,530	7,575	5,476	25,581	13,322	8,238	10,242	31,802
1989	580	654	4,001	5,235	9,759	8,571	4,490	22,820	10,339	9,225	8,491	28,055
1990	628	641	3,855	5,124	11,522	10,064	4,757	26,343	12,150	10,705	8,612	31,467
1991	573	542	3,393	4,508	11,335	8,910	4,521	24,766	11,908	9,452	7,914	29,274
1992	506	423	2,656	3,585	8,517	7,668	3,995	20,180	9,023	8,091	6,651	23,765
1993	485	514	2,514	3,513	8,244	9,350	4,214	21,808	8,729	9,864	6,728	25,321
1994	614	777	2,203	3,594	6,166	8,200	3,070	17,436	6,780	8,977	5,273	21,030
1995	734	835	1,960	3,529	6,144	6,534	2,448	15,126	6,878	7,369	4,408	18,655
1996	822	943	2,180	3,945	7,275	8,412	3,108	18,795	8,097	9,355	5,288	22,740
1997	904	856	2,352	4,112	7,134	9,424	3,656	20,214	8,038	10,280	6,008	24,326

Estimates After Data Correction

Year	Exploratory				Developmental				All Wells			
	Oil	Gas	Dry	Total	Oil	Gas	Dry	Total	Oil	Gas	Dry	Total
1978	1,191	1,792	8,054	11,037	17,874	12,613	8,537	39,024	19,065	14,405	16,591	50,061
1979	1,335	1,920	7,478	10,733	19,368	13,250	8,560	41,178	20,703	15,170	16,038	51,911
1980	1,807	2,126	9,052	12,985	30,784	15,158	11,562	57,504	32,591	17,284	20,614	70,489
1981	2,747	2,574	12,357	17,678	40,821	17,552	15,418	73,791	43,568	20,126	27,775	91,469
1982	2,634	2,217	11,320	16,171	36,495	16,729	14,904	68,128	39,129	18,946	26,224	84,299
1983	2,233	1,715	10,197	14,145	34,869	12,843	13,961	61,673	37,102	14,558	24,158	75,818
1984	2,518	1,682	11,458	15,658	40,021	15,357	14,198	69,576	42,539	17,039	25,656	85,234
1985	2,240	1,495	9,189	12,924	32,691	12,698	11,838	57,227	34,931	14,193	21,027	70,151
1986	2,004	1,156	5,726	8,886	16,974	7,256	6,855	31,085	18,978	8,412	12,581	39,971
1987	1,692	1,095	5,187	7,974	14,402	6,843	5,903	27,148	16,094	7,938	11,090	35,122
1988	1,498	1,253	4,850	7,601	12,063	7,206	5,179	24,448	13,561	8,459	10,029	32,049
1989	1,216	1,502	4,025	6,743	8,967	7,921	4,144	21,032	10,183	9,423	8,169	27,775
1990	1,262	1,527	3,838	6,627	10,919	9,433	4,462	24,814	12,181	10,960	8,300	31,441
1991	1,221	1,247	3,420	5,888	10,500	8,233	4,184	22,917	11,721	9,480	7,604	28,805
1992	995	896	2,616	4,507	7,769	7,265	3,493	18,527	8,764	8,161	6,109	23,034
1993	867	879	2,585	4,331	7,385	9,062	3,639	20,086	8,252	9,941	6,224	24,417
1994	817	987	2,393	4,197	5,772	8,558	2,851	17,181	6,589	9,545	5,244	21,378
1995	855	975	2,285	4,115	6,647	7,277	2,751	16,675	7,502	8,252	5,036	20,790
1996	822	1,020	2,194	4,036	7,042	8,094	2,982	18,118	7,864	9,114	5,176	22,154
1997	801	961	2,056	3,818	7,215	9,229	3,159	19,603	8,016	10,190	5,215	23,421

Sources: Energy Information Administration (EIA). **Estimates Prior to Data Correction:** *Monthly Energy Review (MER)* (August 1997). Estimates for 1997 are double the figure for the first 6 months. **Estimates After Data Correction:** As published in the February 1998 issue of *MER*.

Figure SR4. Exploratory Success Rates for Previous and Revised Data



Source: Energy Information Administration, Office of Oil and Gas.

understanding of the industry that their implications for analysts must be considered.

Implications of the data shift. The significance of the impact is likely to depend on the particular use of the data. EIA itself uses these data in a variety of applications. In particular, they support analysis integral to two prominent information products: the *Annual Energy Outlook* (AEO) and the *Short-Term Energy Outlook*. The drilling data are important measures used in the National Energy Modeling System, which is the tool used to produce the integrated energy market projections that are published in the AEO. Even knowing the role of the data in support of the AEO, the effect on the results is uncertain.

The productivity of exploratory drilling is represented in the NEMS by finding rates, which are measured as the ratio of reserve additions to wells completed. The increase in the revised gas exploratory well counts raise the denominator, so the finding rate will shift downward correspondingly. This reduction in productivity would lower the available market supplies in future years, all else being equal. However, the corresponding analysis that determines the level of drilling must also be revised, because it is predicated on a low well count. Thus, the expected drilling response under varying conditions is underestimated in the projection because it is calibrated to an inaccurately low benchmark. When these functional relations are reestimated, drilling response should be greater in the future than previously expected. The net result of these two offsetting influences on the supply outlook is not obvious.

The long-term supply outlook is also stimulated by improved performance in the search for oil and gas as indicated by the higher success rates, which enhances the economic attractiveness of exploratory drilling opportunities. Dry holes represent an unavoidable part of the search for oil and gas that add to project cost, lowering the expected profitability of exploratory projects. Higher average success rates imply that the drilling costs of an average exploration project are reduced. This improvement to industry economics suggests that the economically recoverable portion of the technically recoverable resource base may be larger than previously estimated. The volumetric impact of this benefit should be limited, however, because this improvement will impact oil and gas accumulations that are marginal in size anyway. The large fields that provide the bulk of expected market supplies are unaffected.

Maintaining data quality in the future. EIA conducts a wide range of data quality activities to monitor its data systems. The long-lived problems affecting the drilling

data series maintained by EIA are a clear indication that further efforts in this regard are necessary. Development of new standards for data quality assurance for this series will proceed along a number of paths including closer monitoring of the as-received well completion data and comparison of estimated drilling data by completion month to other data that are available on a timely basis. An indication of the difficulty in establishing any definitive set of guidelines can be seen in the data for wells reported each month. While the well counts in the majority of months vary by 20 percent or less, fluctuations exceeding 30 percent are common. Even relative shifts beyond 40 percent are frequent enough as not to be necessarily considered a cause for alarm (Figure SR5).

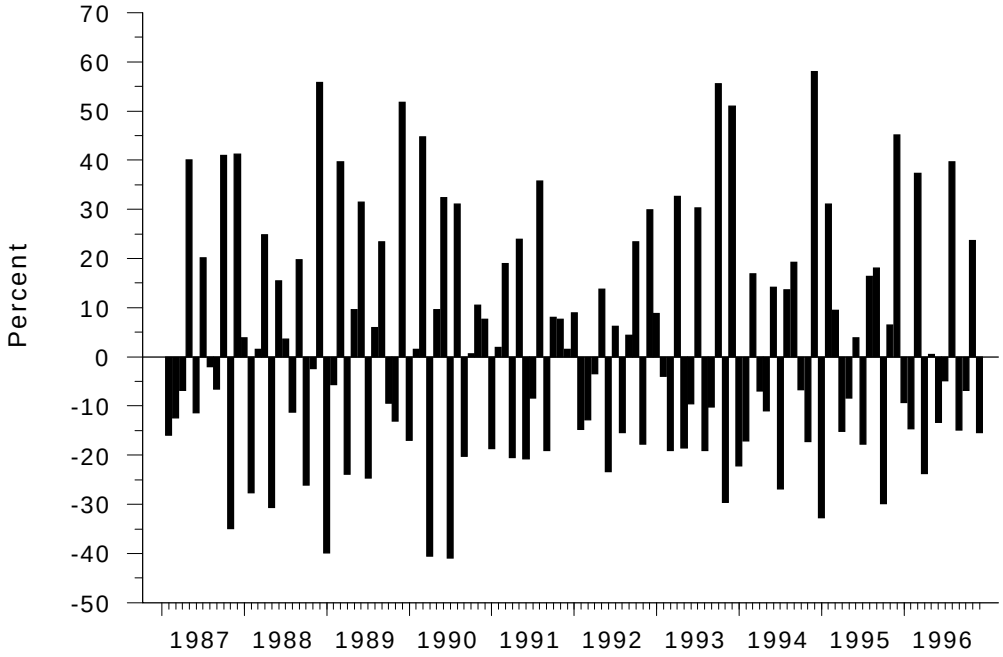
Drilling-related data are used to evaluate the estimates coming from WELCOM. These data include rotary rigs running, seismic crews, and oil and gas prices. EIA staff is also considering the possibility of obtaining relevant data that may exist at the State level in the major producing States to corroborate the EIA estimates. Drilling-related measures from the major States, while not comprehensive, could prove to be useful as a comparator, since most drilling occurs in a limited number of States. These enhanced monitoring actions will avoid future data problems that might require extensive revision of the series.

Conclusions

EIA has inadvertently been publishing erroneous estimates of oil and gas well drilling activity since the late 1980s. Problems in the raw data obtained from a vendor were detected in late 1997 and have since been resolved. Monthly publication of the corrected EIA drilling activity estimates series resumed with the February 1998 edition of the *Monthly Energy Review*. The corrected series' major characteristics are comparable in most regards with the prior erroneous series. There are, however a few exceptions that directly bear on fundamental aspects of the industry such as its responses to changing economic conditions and its approach to drilling opportunities.

EIA does not collect the raw drilling data itself, so some data errors may remain extremely difficult to discover, identify, and remedy in a timely manner. However, EIA will make every possible effort to ensure that the same data problem, or a data problem of similar magnitude, does not happen again.

Figure SR5. Change in Wells Reported Compared to Prior Month



Source: Energy Information Administration, Office of Oil and Gas.