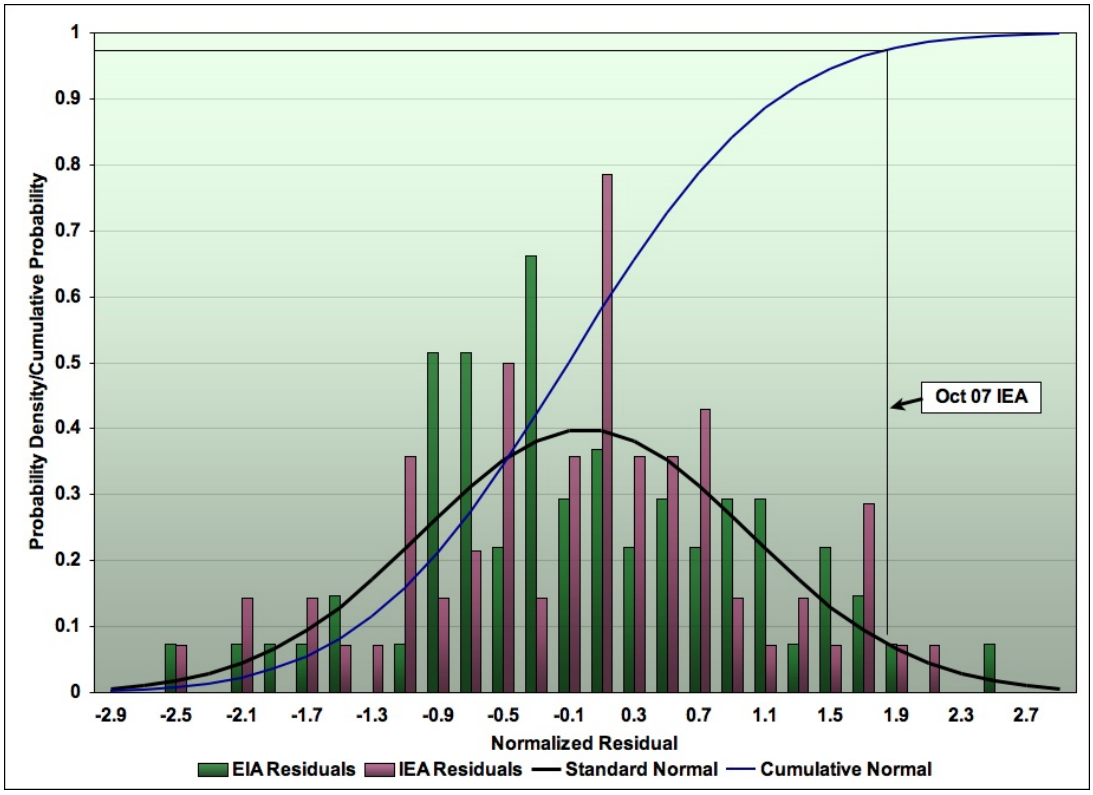
The Oil Drum
DISCUSSIONS ABOUT ENERGY AND OUR FUTURE

Does the latest IEA number matter?

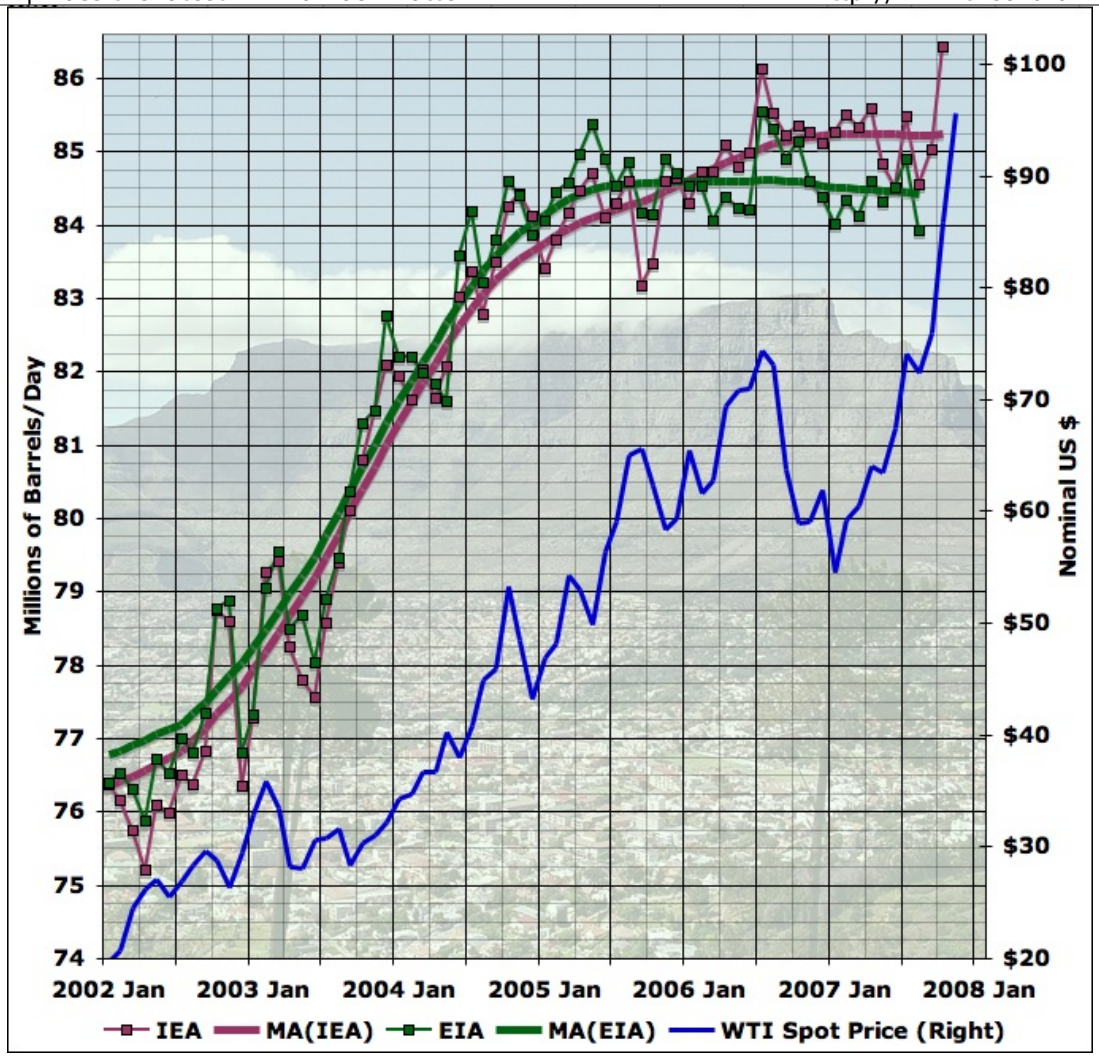
Posted by Stuart Staniford on December 3, 2007 - 11:00am
Topic: Supply/Production
Tags: eia, iea, peak oil, plateau [list all tags]



Normalized histogram of residuals from moving average for EIA and IEA total liquids production, Jan 2002-Oct 2007, together with standard normal distribution, and cumulative probability function for normal. October IEA residual value is emphasized with construction lines and label.

Several people have viewed the October IEA report on total liquid fuel as significant, maybe even a big deal. For those not tracking every month's oil statistics with bated breath, the news was that total fuel production averaged 86.43 million barrels per day, which is higher than the previous high month (July 2006 for this series).

Here's what the data point looks like in the context of the rise and then plateau of the last five years:



Average daily total liquid production, by month, from EIA (green) and IEA (plum), together with 13 month centered moving averages of each line, recursed once (LHS). WTI spot price (blue - RHS). Click to enlarge. Graphs are not zero-scaled. Source: IEA Oil Market Reports, and EIA International Petroleum Monthly Table 1.4. The IEA line is taken from Table 3 of the tables section at the back of the OMR in the last issue for which the number for that month is given. WTI spot price is from the EIA with November estimated from average of daily figures available through the 27th of the month.

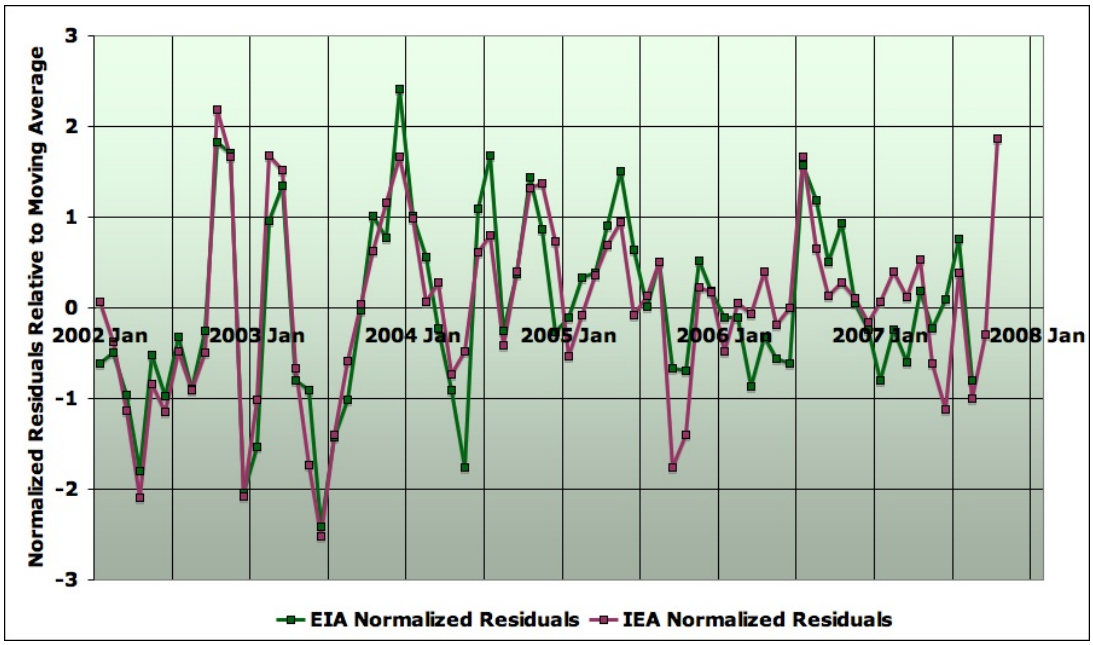
So does this mean that there is "some relief" in oil supplies on the way? Or even that "July 2006 wasn't peak oil after all. Oil production is still growing!"

I would argue that, just as it's premature to declare that oil is already in overall (ie net) decline, so it's also premature to declare that this month's data proves oil supply is growing again. As the graph above should make clear, in addition to the overarching trend of flattening supply, oil production experiences substantial month-to-month fluctuations (aka noise).

Stuff happens. One month a big new project ramps up, another month there's a fire or a storm or a hijacking. Every month there are outages for maintenance and accidents. All this causes production to be more or less each month - some months everything comes together and there is more oil than the average trend would predict. Other months everything falls apart, and there is less. So the question about this most recent month should be whether or not it is a large enough increase to stand out from the background fluctuations that are normal for the globe's oil operations?

And the answer to that appears to be no.

A quick, slightly rough, but I think adequate, way to get at this question is to look at how the monthly production data in the plateau graph above fluctuate around their average trend. If we subtract the moving average of each series from the data, we get what is called the residuals:

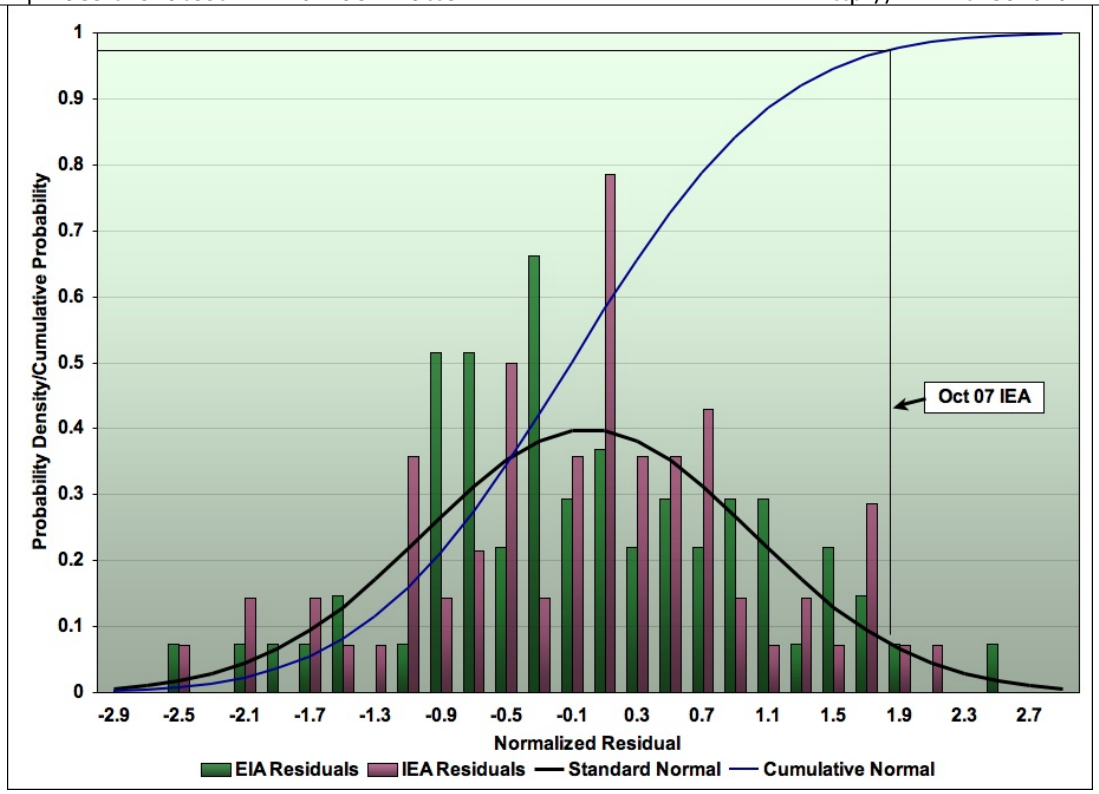


Residuals from moving average (13 month centered moving averages of each line, recursed once) from EIA (green) and IEA (plum), normalized to have a standard deviation of one. Each series is residual to its own moving average. Click to enlarge. Graphs are not zero-scaled. Source: see previous graph.

This is a striking graph. One thing that is interesting is the high degree of agreement between the IEA and the EIA. They seem to agree on the noise better than the trend! Formally, the correlation co-efficient between the two series is $r^2 = 87\%$.

There's also a modest level of auto-correlation in the series - each point is somewhat biassed to be near the previous month's observation, rather than fully independent of it. More on that later.

Another thing to note is that, looked at this way, the IEA's most recent point looks high, but not unprecedentedly so. We can formalize this. Since the noise in global oil production is the sum of many different things all over the world which can cause oil production to be somewhat more or somewhat less, the Central Limit Theorem suggests that the noise will roughly follow the normal distribution. If we make a histogram of both sets of residuals, we can see:



Normalized histogram of residuals from moving average for EIA and IEA total liquids production, together with standard normal distribution, and cumulative probability function for normal. October IEA residual value is emphasized with construction lines and label.

Looks pretty normal. There's a little lumpiness because there's not that much data (70 points or so), and it's a little autocorrelated. There's also a little evidence of too many points too far out in the tails - they may be a little heavier than normal. Still, not bad. I'm willing to treat the noise as normal (which, if anything, will overstate the significance of observations in the tails).

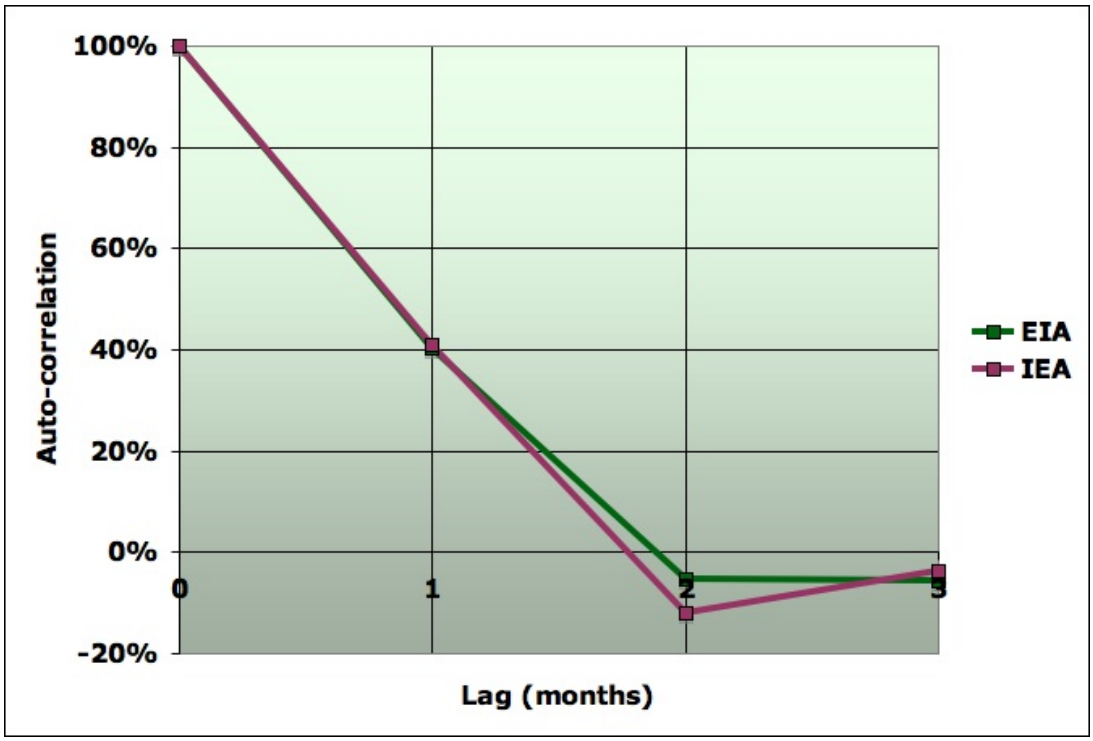
So, you might look at this graph and notice that the October 2007 point is at the 97th percentile which is more than the 95th percentile; isn't this data point a statistically significant departure from trend? (At least if you were convinced oil production had to rise, so you were willing to declare significance on a one-tailed test, and were willing to ignore the slight indication that the tails might be a bit fat for a normal distribution).

I don't think that's the right way of looking at this observation. If we had not had any data till now, and had just obtained this data to do a particular experiment, that would be a valid analysis. However, that's not what's been going on here. At least here at TOD, we've been staring at this monthly data regularly for two years, and debating the meaning of the plateau (and I think the peakoil.com community may even have a few more months of consideration behind it). At any time during that period a rise like this would have seemed significant to some people and led to cries of "the plateau is over!".

So we have to be careful here to take account of what is called data mining bias. This is the general problem that if you look at a dataset long enough, in enough different ways, you can find statistically significant observations just by chance. In our case, if we watched the monthly data for a thousand months, even if it was perfectly flat in trend, we would likely see a 1 in a 1000 jump up in the data at some point just because eventually unlikely things happen if you look long enough.

So, in this case, if we've been looking at the data for 24 months, perhaps the right question is "What is the odds of seeing an observation in the 97th percentile of a standard normal distribution if we try 24 times?" If that number was less than 5%, perhaps that is the criteria for this observation to be significant? Clearly, the odds of hitting the 97th percentile are a lot better with 24 tries than with 1, so this suggests the observation is less significant.

Well, I don't think this is quite right either. The reason is the auto-correlation in the series. The effect of this is to make it as though we have less truly independent data than the number of points in the series. This next graph shows the auto-correlation at lags of 1, 2, and 3 months for both IEA and EIA total liquids residuals.



Auto-correlation for residuals from moving average for EIA and IEA total liquids production, at lags up to 3 months.

As you can see, there is an r^2 of about 40% between one observation and the next, but by two months apart, the auto-correlation is pretty negligible (it's actually very slightly negative).

So roughly speaking, only 60% of the variance in each data point in the residual series is really "new" noise, while 40% is "old" noise left over from the last point. Ditto for the next point, and so on. Thus it's as though we have 0.6*24 ~ 15 independent data points. So, with 15 data points, what is the odds of one of them hitting the 97th percentile? Well, the distribution that tells us the odds of hitting a certain number of things out a lot of independent tries is the binomial distribution. We'd like to know the odds of getting 1 or more thing out of 15 tries at getting something which has a 3% probability per try. We could look it up in a table, but this case can actually be computed from elementary probability. The odds of **not** reaching the 97th percentile on a single try are 0.97 (actually, the odds for our particular residual are 0.9704). The odds of not reaching it on any of 15 tries is 0.9704^{15}. And so the odds or reaching it on one or more tries are $1 - 0.9704^{15} = 36\%$. In other words, watching the IEA data for two years, the chances of seeing a residual this big are more than a third.

Thus, it isn't significant at all, and is not meaningful evidence that anything is going on more than

the usual noise that affects the amount of oil produced each month.

Nor will it be particularly significant if the EIA series also makes a big jump up when we finally see their October number. The 87% r^2 between the two series makes that more likely than not.

Of course, if the IEA series stayed this high for another couple of months, that would be another story altogether. Given the lag-one auto-correlation, a single extra high month wouldn't compel. But three consecutive residuals this high would be very unlikely based on the past history.

And for that reason, my bet is that it won't happen that way. If it **were** to happen, I would accept that as pretty strong evidence that something new was happening and oil production was now increasing again.

Finally, as an aside, there's a slight mystery here. With a one month lag autocorrelation of 40%, I'd have expected a two month lag of around 15%. The fact that both EIA and IEA lag 2 and lag 3 autocorrelations are very slightly negative seems to me rather odd. Anyone got any ideas for a mechanism that would tend to push oil supply slightly in the opposite direction from where it was going, with a lag of about 2 months? OPEC adjusting based on OECD inventories maybe?