Predicting US Production with Gaussians
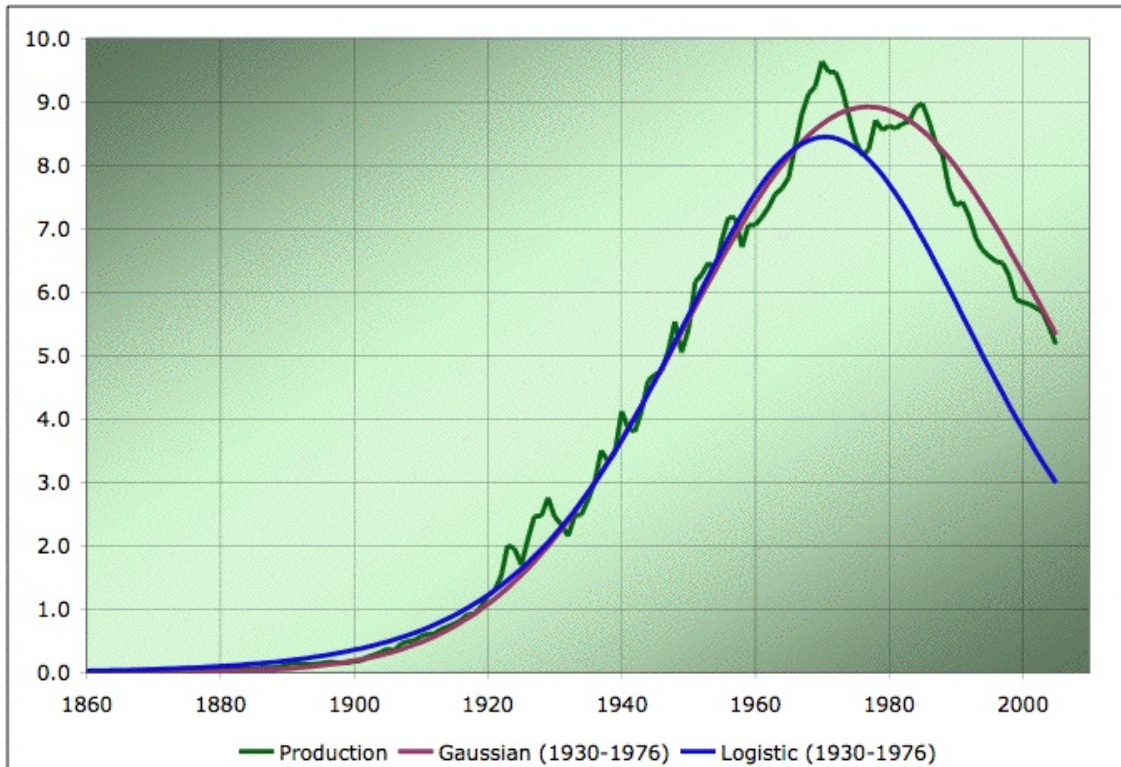
Posted by Stuart Staniford on January 8, 2006 - 4:49am
Topic: Supply/Production
Tags: hubbert peak, oil prices, peak oil, united states [list all tags]

EIA Field production of crude in the US, logistic (Hubbert) fit based only on 1930-1976 data, and Gaussian fit based on same time interval. Source: EIA for the data.

So, I was fooling around tonight, and made a long term graph of US production growth rates (year-on-year), which looks as follows. Because the data are so noisy, I fit a polynomial to it just to get a sense of the trend. The polynomial came out almost a straight line. I varied the degree (in the graph it's a polynomial of degree 6), but it always wanted to be more or less straight.
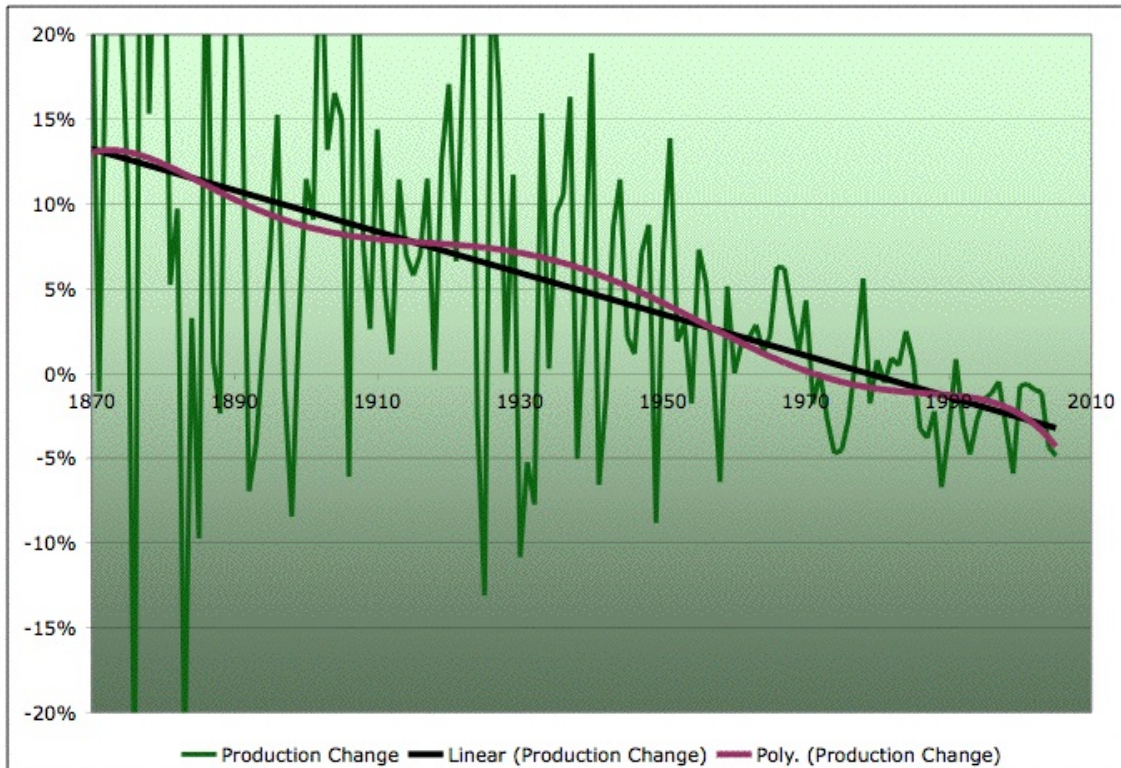
Year-on-year change in EIA Field production of crude in the US, with linear and sixth order polynomial fit. Source: EIA for the data.

That's not what the logistic would say - the logistic would call for an S shaped decline in the growth rate, starting at K (which is around 6%). Of course, we know the logistic is not that great at modeling the early production. Still, that straight line is really sticking out. Hmmm. Scratch head, write a few equations, turns out that the function that has a linearly decreasing growth rate is a Gaussian. I've vaguely heard of people using Gaussian's instead of logistics as models of the peak, but haven't played with it myself before tonight.

So, plot the log of production versus time and fit a parabola: Oh my.

$$y = -0.00052738x^2 + 2.08841650x - 2065.40103703$$
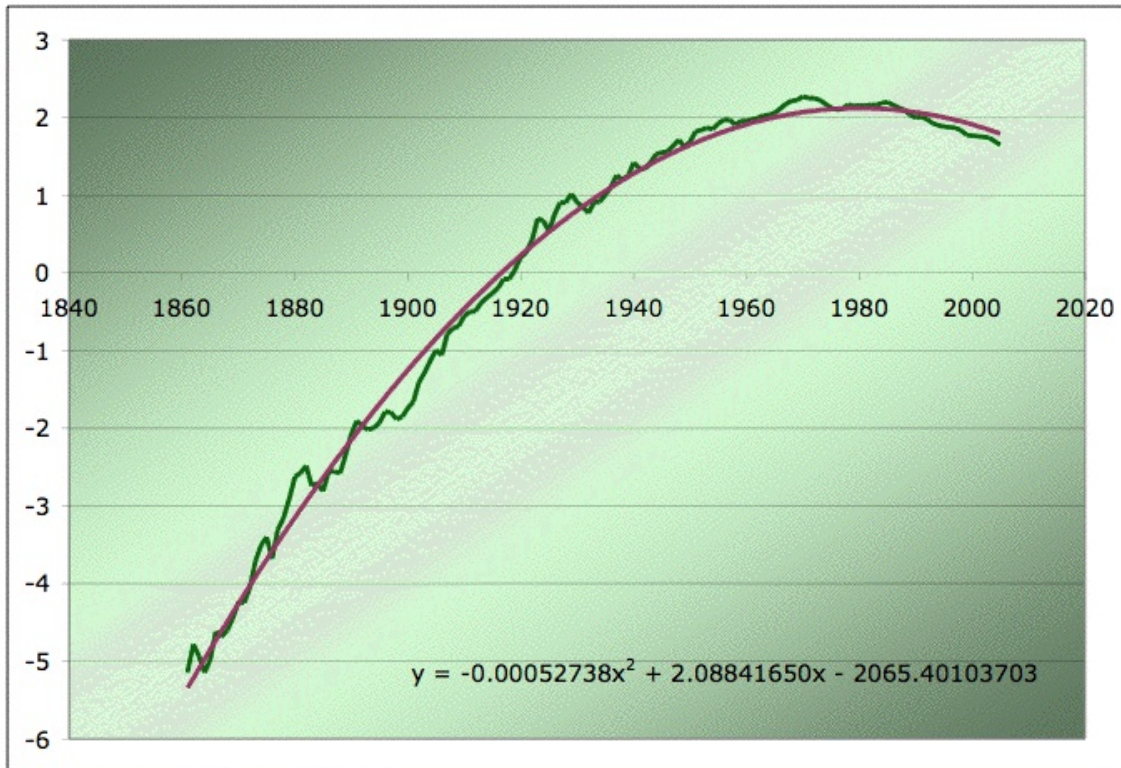
Natural log of EIA Field production of crude in the US, with quadratic fit. Source: EIA for the data.

Pretty good fit across the whole range. There's a very famous theorem in statistics (the central limit theorem) which says roughly that if you add a whole bunch of variables together which are identically distributed, the resulting sum will have a Gaussian distribution. You could argue that something similar is causing this, but the the things being added together are not obviously identically distributed. It's not clear to me why the central limit theorem would apply to a dynamical process in time - the time profile of oil production is not a statistical sampling process, it's an economic/stochastic/sociological spread process through a complex geologic reality. More thought required here.

There must be references on this surely. But I haven't found them in my literature search to date, and can't quickly find them now. Anyone?

Anyway, to get a quick feel for prediction, I repeated the thing I did Thursday night of seeing what would happen if you were to use the model to predict production forward:
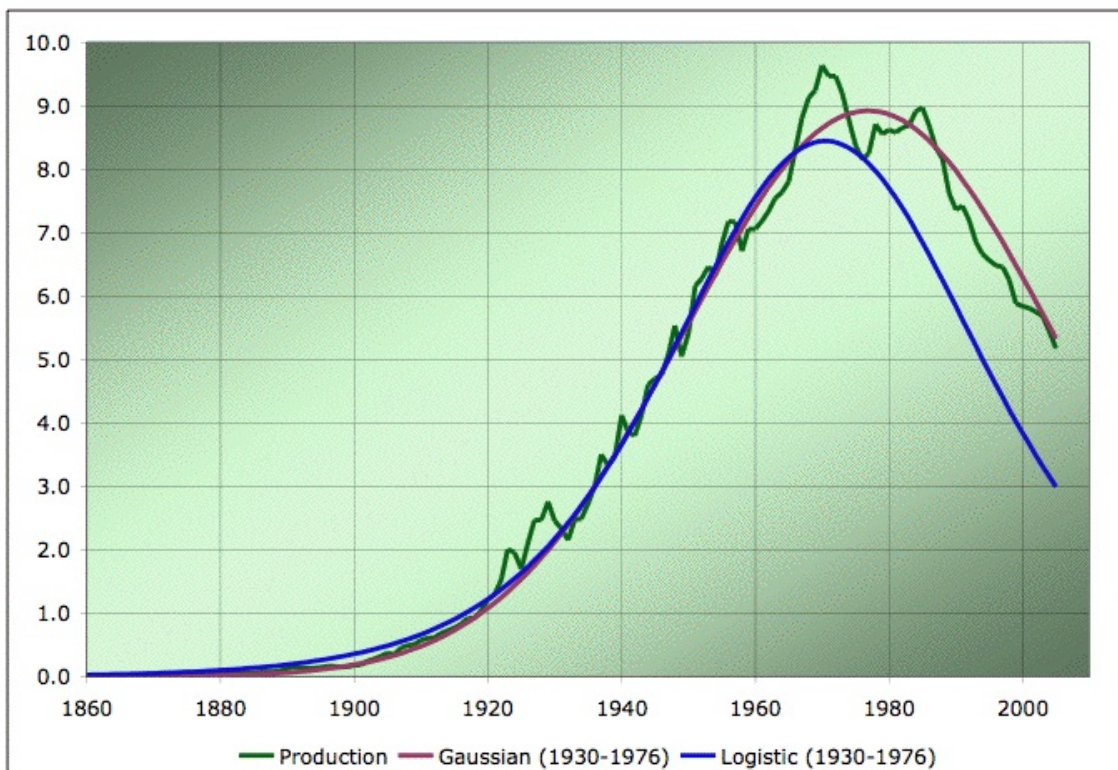
EIA Field production of crude in the US, logistic (Hubbert) fit based only on 1930-1976 data, and Gaussian fit based on same time interval. Source: EIA for the data.

Yikes. That's really good. Not sure if the Gaussian will always do so well, but this is certainly interesting...