Probability in the Engineering and Informational Sciences, **25**, 2011, 205–217.
doi:10.1017/S0269964810000355

OIL PRODUCTION MODELS WITH NORMAL RATE CURVES

Dudley Stark
School of Mathematical Sciences
Queen Mary, University of London
London E1 4NS, United Kingdom
E-mail: D.S.Stark@maths.qmul.ac.uk

The normal curve has been used to fit the rate of both world and US oil production. In this article we give the first theoretical basis for these curve fittings. It is well known that oil field sizes can be modeled by independent samples from a lognormal distribution. We show that when field sizes are lognormally distributed, the starting time of the production of a field is approximately a linear function of the logarithm of its size, and production of a field occurs within a small enough time interval, then the resulting total rate of production is close to being a normal curve.

1. INTRODUCTION

The rate of oil production in the United States was fitted using the logistic curve by Hubbert [7] in 1956. He gave a range of estimated dates for the peak of the rate of oil production from 1965 to 1970, which included the actual peak that occurred around 1970. According to Deffeyes [4], the normal curve gives an even better fit. World oil production is fitted in Deffeyes [4] as well. Hubbert's method of fitting using logistic curves, called the Hubbert linearization method, is described in detail in Deffeyes [4]. Not all curves describing the rate of oil production are bell-shaped, as is the case with logistic and normal curves; see Brandt [3]. Even so, the facts that the logistic curve is the solution of the logistic differential equation and that the normal curve arises from the central limit theorem are tantalizing indications that some mathematical justification for curve fitting by bell-shaped curves should exist.

The *total rate of production* of an area is the sum of the rate of production of its constituent fields. In this article we give conditions under which the total rate of production follows an approximately normal-shaped curve. This supplies the first theoretical framework justifying the curve fitting of the total rate of production by a normal curve. (It happens not to involve the central limit theorem.)

This article proposes a rationale behind fittings of the total rate of production by normal curves. There are two basic assumptions in this article. The first assumption is that the time period that individual fields are in production should be small compared with the width (which remains to be defined) of the total rate of production. The second assumption is that the starting time of production of a field should depend roughly on its size.

Given a particular oil field, its *field size* is the total amount of oil it produces. It is well known that field sizes can be modeled by independent samples from either lognormal or Pareto distributions. The lognormal distribution seems to be better for modeling the discovery process and the Pareto distribution is better for modeling the sizes of all fields in an area; see Barton and Scholz [1]. Under the assumption that field sizes are lognormally distributed, we show that it suffices for the starting time of the production of a field to be approximately a linear function of the logarithm of its size, and for production to occur within a small enough time interval, it suffices for the total rate of production to be close to being a normal curve.

The curves representing the total rate of production generated by our model are not exactly normal but are close to being normal in a way that can be quantified. In order to quantify how close to normal are our total rate of production curves, we use the Kolmogorov–Smirnov distance between our total rate of production curves and their associated normal curves. Actually, we compare each of them with a *limiting rate of production* curve. The smaller the Kolmogorov–Smirnov distance, the closer the limiting rate of production curve is to being normal. This seems to be the first use of a probability metric in studying oil production.

Not much theoretical work has been done on oil production models and no work has been done on models producing bell-shaped curves. Michel [10] used a model in which the shapes of field rate of production curves are constant and rescaled according to the size of the field. He optimized his model to fit North Sea oil production. Stark [11] analyzed explicitly a variant of Michel's model that was first presented in Bentley [2]. Bentley's model does not give realistic total rate of production curves, since they are concave over an initial segment and then convex over the remaining real number line. These models seem to result in total rate of production curves that are roughly Gamma-shaped. In this article the shapes of field rate of production curves are random and the total rate of production curves are close to being normal.

We state our main results in Section 2. Proofs for a simplified model in which fields are produced instantaneously at a time determined by the field size is described in Section 3. In Section 4 we prove a proposition ensuring the existence of a limiting rate of production curve. In Section 5 we prove our results for lognormally distributed field sizes and the approximately normal-shaped limiting rate of production curves. In Section 6 we make some remarks about Pareto-distributed field sizes and the approximately logistic-shaped limiting rate of production curves. In Section 7 we discuss our results and their implications for the study of oil production.

2. MAIN RESULTS

We suppose that field sizes X_i, $i = 1, 2, \ldots, n$, are independent and identically distributed. Let $G_n(t)$ denote the amount of oil produced by these n fields before time t and define $S_n = \sum_{i=1}^{n} X_i$ to be the total amount of oil produced. Note that $S_n = \lim_{t\to\infty} G(t)$. We can turn $G_n(t)$ into a cumulative distribution function by dividing by S_n:

$$F_n(t) = \frac{G_n(t)}{S_n}. \tag{1}$$

The natural interpretation of $F_n(t)$ as a cumulative distribution function (c.d.f.) is that it is the probability that an infinitesimal piece of oil, chosen randomly from all of the oil produced from n fields, is produced before time t.

One of our basic assumptions is that the time period that individual fields are in production should be small compared with the width of the total rate of production. Taking that assumption to its extreme results in a model in which the oil from individual fields is produced instantaneously. The times of production are assumed to be deterministically determined by the field sizes.

In the *instantaneous production model*, we suppose that the oil from field i is produced instantaneously at a deterministic time $\phi(X_i)$, where $\phi : \mathbb{R} \to \mathbb{R}$ is a continuous, monotonically decreasing function (because we expect that the largest fields are discovered and produced first). We therefore have

$$G_n(t) = \sum_{i=1}^{n} X_i I[X_i > \phi^{-1}(t)],$$

where

$$I[X_i > \phi^{-1}(t)] = \begin{cases} 1 & \text{if } X_i > \phi^{-1}(t) \\ 0 & \text{if } X_i \le \phi^{-1}(t). \end{cases}$$

To quantify how close $F_n(t)$ is to being normal, we introduce the Kolmogorov–Smirnov metric on c.d.f.s. Given two c.d.f.s F and $\tilde{F}$, the Kolmogorov–Smirnov distance between them is defined to be

$$d(F, \tilde{F}) = \sup_{t\in\mathbb{R}} \left| F(t) - \tilde{F}(t) \right|.$$

The Kolmogorov–Smirnov distance is used in the Glivenko–Cantelli theorem and in studying the Kolmogorov–Smirnov statistic; see Durrett [5].

Let X be a random variable with the same distribution as the X_i.

LEMMA 1: *Suppose that* $X \ge 0$ *and* $0 < \mathbb{E}(X) < \infty$. *Let*

$$F(t) = \frac{\mathbb{E}\left(XI[X > \phi^{-1}(t)]\right)}{\mathbb{E}(X)}. \tag{2}$$

Then

$$\lim_{n\to\infty} d(F_n, F) = 0 \quad a.s.$$

It can be checked that for any continuous c.d.f. $F(t)$ and continuous random variable X, the function $\phi(x)$ defined by

$$\phi(x) = F^{-1}\left(\frac{\mathbb{E}(XI[X > x])}{\mathbb{E}(X)}\right) \tag{3}$$

satisfies (2).

A lognormally distributed random variable $X \sim \text{Lognormal}(\mu, \sigma^2)$ has the property that

$$\log X \sim \text{N}(\mu, \sigma^2) \tag{4}$$

is normally distributed. A curve $H(t)$ is the c.d.f. of the $\text{N}(m, s^2)$ distribution if

$$\begin{aligned} H(t) &= \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}\,s} \exp\left(-\frac{(x-m)^2}{2s^2}\right) dx \\ &= \Phi\left(\frac{t-m}{s}\right), \end{aligned}$$

where $\Phi(t) = \int_{-\infty}^{t} (1/\sqrt{2\pi}) e^{-x^2/2}\, dx$. When X is lognormally distributed, we have the following result.

Proposition 1: *If $X \sim \text{Lognormal}(\mu, \sigma^2)$ and $\phi(x) = -\alpha \log x + \beta$, $\alpha > 0$, $\beta \in \mathbb{R}$, then $F(t)$ defined by (2) is given by $F(t) = \Phi((t + \alpha\mu + \alpha\sigma^2 - \beta)/\alpha\sigma)$. Thus, $F(t)$ is the c.d.f. of a $\text{N}(\beta - \alpha\mu - \alpha\sigma^2, \alpha^2\sigma^2)$ distributed random variable.*

Remark: It can be shown from (3) that functions $\phi(x)$ of the form in Proposition 1 are the only $\phi(x)$ for which $F(t)$ is the c.d.f. of a normally distributed random variable.

For the instantaneous production model, the derivative $F_n'(t)$ does not exist at all time points and we do not have a total rate of production. We now introduce a general model for which a total rate of production exists and is guaranteed to have a limit. Let $g(t, X_i, Y_i) \geq 0$ be a random function depending on X_i as well as other i.i.d. random elements Y_i that represent auxiliary randomness. We write $\hat{X}_i$ for (X_i, Y_i). The function $g_i(t, \hat{X}_i)$ represents the instantaneous rate of production of field i, so that we require

$$\int_{-\infty}^{\infty} g(t, \hat{X}_i)\, dt = X_i. \tag{5}$$

We now have

$$G_n(t) = \sum_{i=1}^{n} \int_{-\infty}^{t} g(u, \hat{X}_i)\, du. \tag{6}$$

We know that $F_n(t)$ given by (1) and (6) is a c.d.f. because of (1), (5), and (6). It follows from the strong law of large numbers and Fubini's theorem that for fixed

$t \in \mathbb{R}$, $\lim_{n\to\infty} F_n(t) = F(t)$ a.s., where

$$F(t) = \frac{1}{\mathbb{E}(X)} \int_{-\infty}^{t} \mathbb{E}(g(u,\hat{X}))\,du.$$

We define $f_n(t)$ by

$$f_n(t) = \frac{1}{S_n} \sum_{i=1}^{n} g(t,\hat{X}_i), \tag{7}$$

which is the rate of production of the first n fields rescaled according to (1). We define $f(t)$ by

$$f(t) = \frac{\mathbb{E}(g(t,\hat{X}))}{\mathbb{E}(X)}, \tag{8}$$

so that $f_n(t)$ and $f(t)$ satisfy

$$F_n(t) = \int_{-\infty}^{t} f_n(u)\,du \tag{9}$$

and

$$F(t) = \int_{-\infty}^{t} f(u)\,du, \tag{10}$$

respectively. The next proposition shows that under certain conditions, with high probability $f(t)$ is the limiting rate of production curve of $f_n(t)$ as $n \to \infty$.

PROPOSITION 2: *Suppose that* $0 < \mathbb{E}(X)$, *that* $\mathbb{E}(X^2) < \infty$, *that* $\mathbb{E}(g(t,\hat{X})^2) < \infty$ *for all* $t \in \mathbb{R}$, *that* $\mathbb{E}(g(t,\hat{X}))$ *is bounded over compact intervals, and that for some function* $K = K(\hat{X}) > 0$ *such that* $\mathbb{E}(K) < \infty$, *we have*

$$\mathbb{P}\left(\left|g(t,\hat{X}) - g(s,\hat{X})\right| \le K|t-s|\ \forall s,t \in \mathbb{R}\right) = 1. \tag{11}$$

Then, for any $\underline{t} < \bar{t}$, *we have*

$$\sup_{\underline{t}\le t\le \bar{t}} |f_n(t) - f(t)| \xrightarrow{P} 0 \quad \text{as } n \to \infty. \tag{12}$$

Remark: Condition (11) says that $g(t,\hat{X})$ is almost surely Lipschitz with constant $K(\hat{X})$.

Recall that the support of a function $f : \mathbb{R} \to \mathbb{R}$ is defined by support$(f) = \{t : f(t) \neq 0\}$. In the next theorem, our main result, we let $X \sim$ Lognormal(μ,σ^2), so that the conditions on $\mathbb{E}(X)$ and $\mathbb{E}(X^2)$ in Proposition 2 are automatically satisfied, and use the function $\phi(x)$ and the normal curve appearing in Proposition 1. We impose condition (13) on the random subset of $\mathbb{R}$ defined by support$(g(\cdot,\hat{X}))$, which condition implies concentrated production, as opposed to instantaneous production. We therefore call the model in Theorem 1 the *concentrated production model*.

THEOREM 1: *Suppose that* $X \sim \text{Lognormal}(\mu, \sigma^2)$, *that* $\mathbb{E}(g(t, \hat{X})^2) < \infty$ *for all* t, *that* $\mathbb{E}(g(t, \hat{X}))$ *is bounded on compact intervals, and that (*11*) is satisfied, so that by Proposition* 2 *a limiting rate curve* $f(t)$, *defined by* (8), *exists in the sense of* (12). *Let* $\phi(x) = -\alpha \log x + \beta$, $\alpha > 0$, $\beta \in \mathbb{R}$, *and suppose that for some* $\lambda > 0$,

$$\mathbb{P}(\text{support}(g(\cdot, \hat{X})) \subseteq [\phi(X) - \lambda, \phi(X) + \lambda]) = 1. \tag{13}$$

If $H(t)$ *is defined by*

$$H(t) = \Phi((t + \alpha\mu + \alpha\sigma^2 - \beta)/\alpha\sigma),$$

then

$$\limsup_{n \to \infty} d(F_n, H) \leq \frac{\lambda}{\sqrt{2\pi}\alpha\sigma} \quad \text{a.s.}$$

Remark: Condition (13) says that all of the oil produced from field i is produced within the production window λ of $\phi(X_i)$. Therefore, λ is a measure of the weakness of the condition. The width of the limiting rate of production is roughly $\alpha\sigma$, as is evidenced by Proposition 1. Thus, Theorem 1 says that if the production window λ is small compared to the width of the limiting rate of production, then the limiting rate of production is close to being normal. It follows that λ can be large and $d(F_n, H)$ still becomes small as $n \to \infty$.

Remark: Under condition (13) it is natural to assume that $g(t, \hat{X}) = O(X)$ for all t and that $K(\hat{X}) = O(X)$; these assumptions together with $\mathbb{E}(X^2) < \infty$ automatically gives $\mathbb{E}(g(t, \hat{X}))^2) < \infty$ and $\mathbb{E}(K) < \infty$.

3. INSTANTANEOUS PRODUCTION

In this section we prove Lemma 1 and Proposition 1.

PROOF OF LEMMA 1: For each fixed $t \in \mathbb{R}$, by the strong law of large numbers we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mathbb{E}(X) \quad \text{a.s.}$$

and

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i I[X_i > \phi^{-1}(t)] = \mathbb{E}(XI[X > \phi^{-1}(t)]) \quad \text{a.s.}$$

It follows that

$$\lim_{n \to \infty} F_n(t) = F(t) \quad \text{a.s.}$$

By our assumptions, we know that $\phi^{-1}(t)$ is continuous. It follows from our assumption $\mathbb{E}(X) < \infty$ and dominated convergence that the function $\mathbb{E}(XI[X > x])$ is a continuous function of x. Therefore, $\mathbb{E}(XI[X > \phi^{-1}(t)])$ and hence $F(t)$ are continuous functions of t. The remaining part of the proof is exactly like part of the proof of the Glivenko–Cantelli theorem in Durrett [5]. ■

PROOF OF PROPOSITION 1: The expected size of a single field is $\mathbb{E}(X) = e^{\mu+\sigma^2/2}$. Moreover, as $X = e^{\mu+\sigma Z}$, where $Z \sim \mathrm{N}(0,1)$, we have

$$\begin{aligned}
\mathbb{E}\left(XI[X > \phi^{-1}(t)]\right) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} e^{\mu+\sigma y} I[-\alpha(\mu+\sigma y) + \beta \leq t]\, dy \\
&= \int_{(\beta-t-\alpha\mu)/\alpha\sigma}^{\infty} e^{\mu} \frac{1}{\sqrt{2\pi}} e^{-y^2/2+\sigma y}\, dy \\
&= e^{\mu+\sigma^2/2} \int_{(\beta-t-\alpha\mu)/\alpha\sigma-\sigma}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \\
&= \mathbb{E}(X)(\Phi((t+\alpha\mu-\beta+\alpha\sigma^2)/\alpha\sigma). \qquad \mathbf{(14)}
\end{aligned}$$

Using (14) in (2) completes the proof. ■

4. THE LIMITING RATE OF PRODUCTION

In this section we prove Proposition 2.

PROOF OF PROPOSITION 2: Fix $\epsilon > 0$ and choose $\underline{t} = t_0 < t_1 < \cdots < t_L = \bar{t}$ such that

$$\max_{1\leq j\leq L} (t_j - t_{j-1}) < \frac{\epsilon}{3\mathbb{E}(K)}.$$

By the L^2 weak law of large numbers, we can choose $N = N(\epsilon) > 0$ large enough so that

$$\mathbb{E}\left(\left|\frac{1}{n}\sum_{i=1}^{n} g(t_j, \hat{X}_i) - \mathbb{E}(g(t_j, \hat{X}))\right|\right) < \frac{\epsilon}{3L} \qquad \mathbf{(15)}$$

for all $1 \leq j \leq L$ and $n > N$.

Now, let $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_L$ partition $[\underline{t}, \bar{t}]$ in such a way that $t \in \mathcal{A}_j \Rightarrow |t - t_j| < \epsilon/3\mathbb{E}(K)$ for all $1 \leq j \leq L$. Observe that

$$\begin{aligned}
\left|\frac{1}{n}\sum_{i=1}^{n} g(t, \hat{X}_i) - \frac{1}{n}\sum_{i=1}^{n} g(s, \hat{X}_i)\right| &\leq \frac{1}{n}\sum_{i=1}^{n}\left|g(t, \hat{X}_i) - g(s, \hat{X}_i)\right| \\
&\leq \frac{|t-s|}{n}\sum_{i=1}^{n} K(\hat{X}_i) \qquad \mathbf{(16)}
\end{aligned}$$

and that

$$\begin{aligned}\left|\mathbb{E}(g(t,\hat{X}_i)) - \mathbb{E}(g(s,\hat{X}_i))\right| &\le \mathbb{E}\left(\left|g(t,\hat{X}_i) - g(s,\hat{X}_i)\right|\right)\\ &\le \mathbb{E}\left(K(\hat{X}_i)|t-s|\right)\\ &= \mathbb{E}(K)|t-s|. \end{aligned} \tag{17}$$

For each $1 \le j \le L$ and $t \in \mathcal{A}_j$, we estimate

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^n g(t,\hat{X}_i) - \mathbb{E}(g(t,\hat{X}))\right|\\ &\quad\le \left|\frac{1}{n}\sum_{i=1}^n g(t,\hat{X}_i) - \frac{1}{n}\sum_{i=1}^n g(t_j,\hat{X}_i)\right|\\ &\qquad + \left|\frac{1}{n}\sum_{i=1}^n g(t_j,\hat{X}_i) - \mathbb{E}(g(t_j,\hat{X}))\right| + \left|\mathbb{E}(g(t_j,\hat{X})) - \mathbb{E}(g(t,\hat{X}))\right|\\ &\quad\le \left|\frac{1}{n}\sum_{i=1}^n g(t_j,\hat{X}_i) - \mathbb{E}(g(t_j,\hat{X}))\right| + \frac{\epsilon}{3\mathbb{E}(K)n}\sum_{i=1}^n K(\hat{X}_i) + \frac{\epsilon}{3}\end{aligned}$$

by (16) and (17). Therefore, we obtain

$$\begin{aligned}&\sup_{\underline{t}\le t\le \bar{t}}\left|\frac{1}{n}\sum_{i=1}^n g(t,\hat{X}_i) - \mathbb{E}(g(t,\hat{X}))\right|\\ &\quad= \max_{1\le j\le L}\left\{\sup_{t\in\mathcal{A}_j}\left|\frac{1}{n}\sum_{i=1}^n g(t,\hat{X}_i) - \mathbb{E}(g(t,\hat{X}))\right|\right\}\\ &\quad\le \max_{1\le j\le L}\left\{\left|\frac{1}{n}\sum_{i=1}^n g(t_j,\hat{X}_i) - \mathbb{E}(g(t_j,\hat{X}))\right|\right\}\\ &\quad= + \frac{\epsilon}{3\mathbb{E}(K)n}\sum_{i=1}^n K(\hat{X}_i) + \frac{\epsilon}{3}\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}\left(\sup_{\underline{t}\le t\le \bar{t}}\left|\frac{1}{n}\sum_{i=1}^n g(t,\hat{X}_i) - \mathbb{E}(g(t,\hat{X}))\right|\right)\\ &\le \sum_{j=1}^L \mathbb{E}\left(\left|\frac{1}{n}\sum_{i=1}^n g(t_j,\hat{X}_i) - \mathbb{E}(g(t_j,\hat{X}))\right|\right) + \frac{2\epsilon}{3}\\ &\le \epsilon\end{aligned}$$

for $n > N$, where we have used (15) in the last step. As $\epsilon > 0$ is arbitrary, we infer that

$$\lim_{n\to\infty} \mathbb{E}\left(\sup_{\underline{t}\le t\le \bar{t}} \left|\frac{1}{n}\sum_{i=1}^{n} g(t,\hat{X}_i) - \mathbb{E}(g(t,\hat{X}))\right|\right) = 0$$

and hence

$$\sup_{\underline{t}\le t\le \bar{t}} \left|\frac{1}{n}\sum_{i=1}^{n} g(t,\hat{X}_i) - \mathbb{E}(g(t,\hat{X}))\right| \xrightarrow{P} 0. \tag{18}$$

Finally, we observe from (7) and (8) that

$$\begin{aligned}\sup_{\underline{t}\le t\le \bar{t}} |f_n(t) - f(t)| &= \sup_{\underline{t}\le t\le \bar{t}} \left|\frac{1}{S_n}\sum_{i=1}^{n} g(t,\hat{X}_i) - \frac{\mathbb{E}(g(t,\hat{X}))}{\mathbb{E}(X)}\right| \\ &\le \frac{n}{S_n}\sup_{\underline{t}\le t\le \bar{t}} \left|\frac{1}{n}\sum_{i=1}^{n} g(t,\hat{X}_i) - \mathbb{E}(g(t,\hat{X}))\right| \\ &\quad + \sup_{\underline{t}\le t\le \bar{t}} \mathbb{E}(g(t,\hat{X}))\left|\frac{n}{S_n} - \frac{1}{\mathbb{E}(X)}\right| \\ &\xrightarrow{P} 0,\end{aligned}$$

where we have used (18), the assumption that $\mathbb{E}(g(t,\hat{X}))$ is bounded on compact intervals, and the L^2 weak law of large numbers, implying

$$\frac{n}{S_n} \xrightarrow{P} \frac{1}{\mathbb{E}(X)}$$

for the last step. ■

5. CONCENTRATED PRODUCTION

In this section we prove Theorem 1.

PROOF OF THEOREM 1: Because Kolmogorov–Smirnov distance is a metric, we have

$$d(F_n,H) \le d(F_n,F) + d(F,H).$$

It follows from (8), (10), and the assumption that $\mathbb{E}(g(t,\hat{X}))$ is bounded on compact intervals that $F(t)$ is continuous. Hence, as in the proof of Proposition 1, we know that

$$\lim_{n\to\infty} d(F_n,F) = 0 \quad \text{a.s.} \tag{19}$$

It remains to be shown that

$$d(F,H) \leq \frac{\lambda}{\sqrt{2\pi}\alpha\sigma}.$$

By (5), (7), (9), and (13), we have

$$\begin{aligned} F_n(t) &= \frac{1}{S_n}\sum_{i=1}^{n}\int_{-\infty}^{t} g(u,\hat{X}_i)\,du \\ &\geq \frac{1}{S_n}\sum_{i=1}^{n} I\left[X_i > \phi^{-1}(t-\lambda)\right]\int_{-\infty}^{t} g(u,\hat{X}_i)\,du \\ &= \frac{1}{S_n}\sum_{i=1}^{n} I\left[X_i > \phi^{-1}(t-\lambda)\right]X_i. \end{aligned} \tag{20}$$

Together with (19) and the strong law of large numbers, (20) implies

$$F(t) \geq \frac{\mathbb{E}\left(XI[X > \phi^{-1}(t-\lambda)]\right)}{\mathbb{E}(X)}.$$

Now, by a calculation similar to one in the proof of Proposition 1,

$$\begin{aligned} &\mathbb{E}\bigg(XI[X > \phi^{-1}(t-\lambda)]\bigg) \\ &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}e^{\mu+\sigma y}I[-\alpha(\mu+\sigma y)+\beta \leq t-\lambda]\,dy \\ &= \mathbb{E}(X)\Phi((t-\lambda+\alpha\mu-\beta+\alpha\sigma^2)/\alpha\sigma) \\ &= \mathbb{E}(X)\big(\Phi((t+\alpha\mu-\beta+\alpha\sigma^2)/\alpha\sigma)+\Delta_1\big) \\ &= \mathbb{E}(X)(H(t)+\Delta_1)\,, \end{aligned}$$

where

$$|\Delta_1| \leq \frac{\lambda}{\sqrt{2\pi}\alpha\sigma}$$

uniformly for all t.

We also have

$$\begin{aligned} F_n(t) &= \frac{1}{S_n}\sum_{i=1}^{n} I\left[X_i > \phi^{-1}(t+\lambda)\right]\int_{-\infty}^{t} g(u,\hat{X}_i)\,du \\ &\leq \frac{1}{S_n}\sum_{i=1}^{n} I\left[X_i > \phi^{-1}(t+\lambda)\right]X_i, \end{aligned} \tag{21}$$

which results in

$$F(t) \le \frac{\mathbb{E}\left(XI[X > \phi^{-1}(t+\lambda)]\right)}{\mathbb{E}(X)}.$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\Big(XI[X > \phi^{-1}(t+\lambda)]\Big) \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} e^{\mu+\sigma y} I[-\alpha(\mu+\sigma y)+\beta \le t+\lambda]\, dy \\
&= \mathbb{E}(X)\Phi((t+\lambda+\alpha\mu-\beta+\alpha\sigma^2)/\alpha\sigma) \\
&= \mathbb{E}(X)\big(\Phi((t+\alpha\mu-\beta+\alpha\sigma^2)/\alpha\sigma)+\Delta_2\big) \\
&= \mathbb{E}(X)(H(t)+\Delta_2)\,,
\end{aligned}$$

where

$$|\Delta_2| \le \frac{\lambda}{\sqrt{2\pi}\alpha\sigma}$$

uniformly for all t.

The previous considerations show that

$$\Delta_1 \le F(t) - H(t) \le \Delta_2,$$

completing the proof. ■

6. PARETO-DISTRIBUTED FIELD SIZES

In this section we make some remarks about Pareto-distributed field sizes and approximately logistic-shaped total rate of production curves.

A Pareto-distributed random variable $X \sim \text{Pareto}(k, a)$ with parameters $k > 0$, $a > 0$, is defined to have probability density function

$$f_X(x) = \begin{cases} \dfrac{a}{k}\left(\dfrac{x}{k}\right)^{-a-1} & \text{if } x > k \\ 0 & \text{if } x \le k. \end{cases}$$

Mandelbrot [9] plotted cumulative Paretian graphs for oil fields with a between 1.5 and 2.

A logistically distributed random variable $Y \sim \text{Logistic}(\mu, s)$ is defined to have probability density function

$$f_Y(y) = \frac{e^{-(y-\mu)/s}}{s(1+e^{-(y-\mu)/s})^2}. \tag{22}$$

If $X \sim \text{Pareto}(k, 1)$, then it is easily shown that

$$\log(X - k) \sim \text{Logistic}(\log k, 1),$$

which is quite similar to (4).

It can be checked that when $X \sim \text{Pareto}(k, a)$, $a > 1$, and $F(t) = \int_{-\infty}^{t} f_Y(y)\, dy$ with $f_Y(y)$ given by (22) that (3) produces

$$\phi(x) = \mu - s \log\big((x/k)^{a-1} - 1\big),$$

which is close to being linear in $\log x$. Recall that $\phi(x)$ is exactly linear in $\log x$ in Proposition 1.

We cannot generally apply Proposition 2 to Pareto-distributed X_i because the Pareto distribution does not always have a second moment unless $a > 2$. We are not sure whether the second moment condition of Proposition 2 can be weakened.

Another approach that could be taken would be to consider the truncated Pareto distribution, which has been found by Goldberg [6] to fit the size distribution of oil fields. An upper truncated Pareto distribution with parameters $0 < k < v$ and $a > 0$ is defined to have probability density function

$$f_X(x) = \begin{cases} \dfrac{ak^a}{x^{a+1}} \left[1 - (v/k)^{-a}\right]^{-1} & \text{if } k < x < v \\ 0 & \text{if } x \le k \text{ or } x > v. \end{cases}$$

As it has all moments, Proposition 2 could be applied to it directly. We decline to do that here.

7. DISCUSSION

We have analyzed in Theorem 1 the concentrated production model of oil production with approximately normal limiting rate curves. There are several comments that can be made about this model.

1. The model could be taken as applying only to fields below a certain threshold size. For example, the largest field in the United States is Prudhoe Bay, which started production later than one would expect from its size as a consequence of its location in the arctic. Prudhoe Bay skews the total rate of production curve of the United States to the right—see Figure 3 of Laherrere [8]—although not enough to move the peak total rate of production away from 1970. On the other hand, if a few smaller fields were produced earlier than one would expect from their sizes, they would not skew the total rate of production curve by much.
2. Condition (13) probably does not need to hold precisely for the approximation of Theorem 1 to hold. What is intuitively important is that, with high probability $g(t, X)$ is small when t is far from $\phi(X)$.

3. It would be most interesting to have empirical evidence as to whether this model describes actual oil production in some area or areas. This should hold when the production of fields of size x is centered about $\phi(x) = -\alpha \log x + \beta$ for some constants $\alpha > 0$ and $\beta \in \mathbb{R}$. If such evidence exists, then it would be desirable to find some explanation for the appearance of the function $\phi(x)$.
4. The total rate of production curve is approximately normal for US and world oil production. For many other areas, such as the North Sea, it is not; see Michel [10]. This could be because the larger areas have a wider range of field sizes and because their total rate of production curves have greater width compared to the time intervals that individual fields are in production.
5. The function $\phi(x) = -\alpha \log x + \beta$, $\alpha > 0$, $\beta \in \mathbb{R}$, implies that the smallest fields will be produced indefinitely far into the future; any reasonable model producing a bell-shaped curve would have this feature. This is probably an unrealistic assumption. Could this be part of the reason that the current rate of oil production in the United States is higher than was predicted by Hubbert [7] (see Figure 3 of Laherrere [8])? Another reason could be delayed production from large fields such as Prudhoe Bay.
6. We have investigated lognormally distributed field sizes together with normal total production curves for the reason that the analysis seemed natural and elegant. Other combinations of field size distributions and classes of total production curves could also be studied—for example, the ones in Section 6.

References

1. Barton, C.C. & Scholz, C.H. (1995). The fractal size and spatial distribution of hydrocarbon accumulations: Implications for resource assessment and exploration strategy. In C.C. Barton & P.R. La Pointe (eds.), *Fractals in petroleum geology and earth processes*. New York: Plenum Press, pp. 13–34.
2. Bentley, R.W. (2002). Global oil and gas depletion: an overview. *Energy Policy* 30: 189–205.
3. Brandt, A.R. (2007). Testing Hubbert. *Energy Policy* 35: 3074–3088.
4. Deffeyes, K.A. (2001). *Hubbert's peak*. Princeton, NJ: Princeton University Press.
5. Durrett, R. (1991). *Probability: Theory and examples*. New York: Wadsworth & Brooks/Cole.
6. Goldberg, G. (1967). *The Pareto law and the pyramid distribution*. Publication 505. Houston, TX: Shell Development.
7. Hubbert, M.K. (1956). Nuclear energy and the fossil fuels. In *American Petroleum Institute Drilling and Production Practise, Proceedings of Spring Meeting*, San Antonio, TX.
8. Laherrere, J. (2005). Forecasting production from discovery. Talk delivered at the Association for the Study of Peak Oil & Gas (ASPO) Conference, Lisbon, Portugal. Available from www.cge.uevora.pt/aspo2005/abscom/ASPO2005_Laherrere.pdf.
9. Mandelbrot, B.B. (1962). The statistics of natural resources and the law of Pareto. In C.C. Barton & P.R. La Pointe (eds.), *Fractals in petroleum geology and earth processes*. New York: Plenum Press, pp. 1–12.
10. Michel, B. (2010). *Oil production: A probabilistic model of the Hubbert curve*. doi:10.1002/asmb.851.
11. Stark, D. (2008). Peak production in an oil depletion model with triangular field profiles. *Journal of Interdisciplinary Mathematics* 11: 695–706.