# Not so long tail of movies

Kalevi Kilkki
Principal Scientist
Nokia Siemens Networks

4th July 2007

This is an unpublished essay about the specific properties of movie business, particularly in US markets. The paper includes a brief documentation about an addition to the long tail model for cases in which the tail is abruptly interrupted. The original long tail model was introduced in the article "A practical model for analyzing long tails" available at http://www.firstmonday.org/issues/issue12_5/kilkki/. This essay is available at http://kilkki.net/3.

The original long tail formula is the following:

$$F(x) = \frac{\beta}{\left(\frac{N_{50}}{x}\right)^{\alpha} + 1}$$

Where $F(x)$ = the share of total volume covered by objects up to rank $x$

$N_{50}$ = the number of objects that cover half of the whole volume

$\alpha$ = the factor that defines the form of the function

$\beta$ = total volume

As the long the article showed, the model is applicable to diverse cases from books and songs to names and companies. However, there are cases in which the long tail distribution works very well with the most popular objects but does not work at all with less popular objects. One of the clearest cases with this kind of behavior is movies presented in movie theaters. Figure 1 shows the box office sales in US in 2006. The original long tail function is able to accurately describe the sales distribution up to rank of 100, but above that limit the real curve diverges from the long tail distribution. The obvious question is whether we can find a simple function that is able to capture the quite regular behavior of the real distribution. After some experiments I ended up with the following correction function:

$$g(x) = \frac{1}{\left(\frac{M_{50}}{x}\right)^{\gamma} + 1}$$

As can be seen, the correction function is of pretty similar with the original long tail function. Note, however, that here the function is applied to the (probability) distribution itself, not to the cumulative volume. Consequently, the predicted volume of the object with rank x is:

$$f^{*}(x) = \frac{F(x) - F(x-1)}{\left(\frac{M_{50}}{x}\right)^{\gamma} + 1}$$

where F(x) is the original cumulative long tail function, and F(0) = 0.

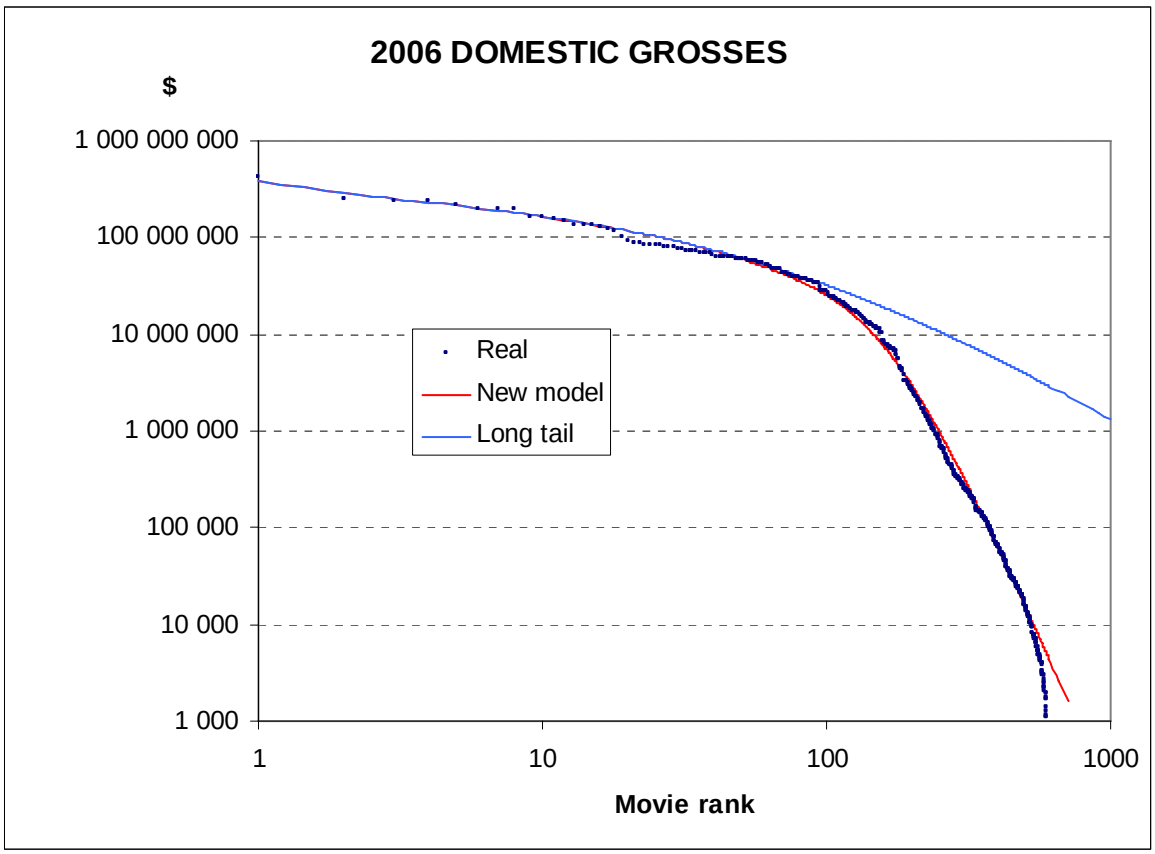



Figure 1. Domestic box office sales in US, 2006.

Now the question is whether this model is able to describe not only the sales of one year and in one market, but also the sales in other years and in other markets. In the next pages there are figures for the US sales from 2002 to 2005, and 2006 sales for France, Germany, UK, Japan and Russia. Obviously, the behavior is very similar in all cases.

A fundamental assumption of the long tail model is that the distributions related to a specific topic, like movies or books, should always be similar. We may, therefore, assume that form factors of both the original function ($\alpha$) and the correction function ($\gamma$) should be constant. The cases used in this brief study indicate that this assumption is valid. A selection of parameters

$\alpha = 0.83$
$\gamma = 5.21$

works in all cases properly as can be seen in the following figures.

In contrast, the other parameters $N_{50}$, $\beta$, $M_{50}$ seem to depend both on the market and to some extent on the year. The optimal parameters are presented in the following table.

Table 1. Model parameters for different years and markets

| | US2006 | US2005 | US2004 | US2003 | US2002 | France | Germany | UK | Japan | Russia |
|---|---|---|---|---|---|---|---|---|---|---|
| $N_{50}$ | 83 | 77 | 61 | 78 | 66 | 39 | 26 | 30 | 31 | 26 |
| $\beta$ | 1.70 | 1.70 | 1.58 | 1.74 | 1.58 | 1.28 | 1.27 | 1.25 | 1.37 | 1.25 |
| $M_{50}$ | 139 | 128 | 132 | 121 | 134 | 215 | 148 | 192 | 111 | 163 |

Now the key claim of this paper is that parameter $\beta$ gives estimation about lost business opportunity because of the some "artificial" constraints in the business ecosystem.

According to these figures the lost business opportunity is 60 - 70% of the current volume in US as $\beta$ varies between 1.58 and 1.74. On the contrary, in Europe the same figure is only 25-30 % (UK, France,

Germany, and Russian). Parameter $M_{50}$ (that defines the threshold for movie rank below which the sales starts to drop rapidly) is about 130 in US, while in France and UK it is clearly higher even though the total volume of the movie business is much smaller.

Thought the numbers seem consistent and the fitting between reality and the modified long tail function is good, we have to be careful when making any conclusions. There are two basic questions: First, what is the reason for the specific long tail behavior in case of movies? Secondly, what would happen if the business model is somehow changed?

Here are some tentative thoughts about these questions. It should be stressed that much more profound studies are needed in order to give certain answers to any of the questions.

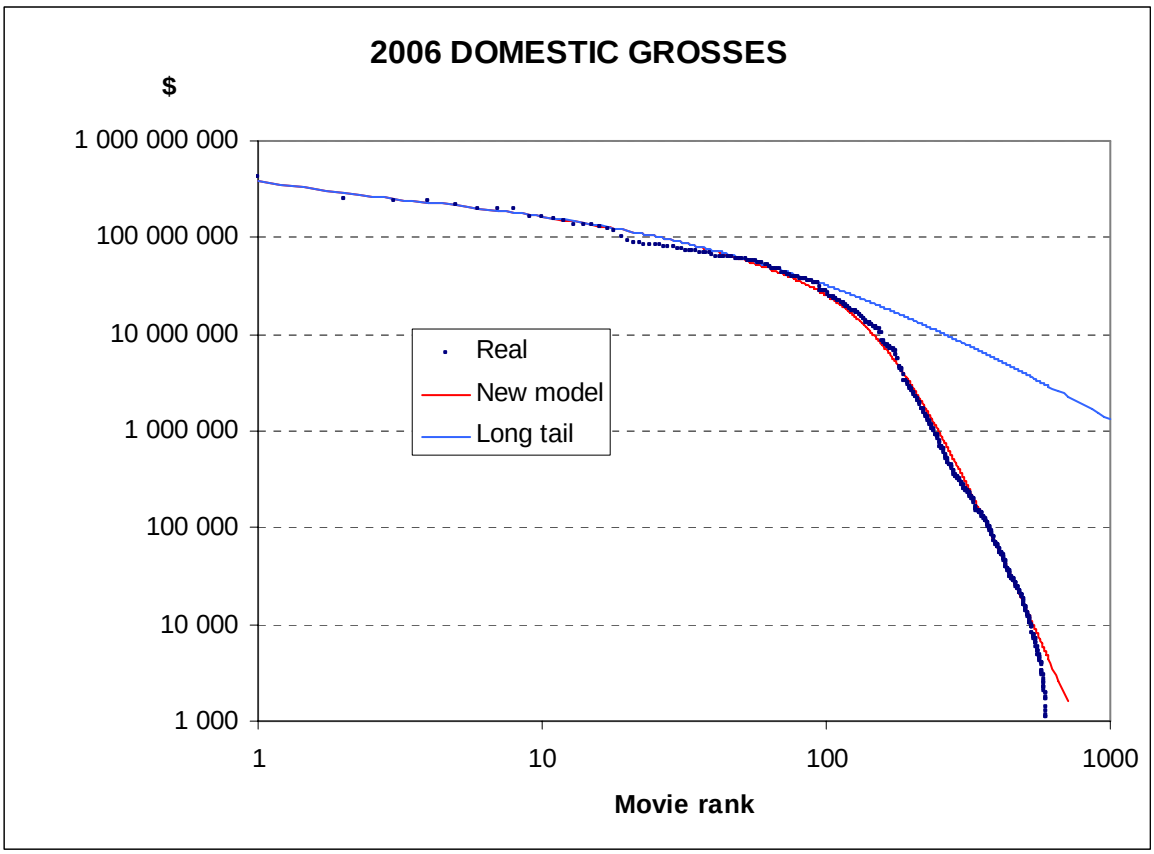
First about the possible reasons. My assumption is that the main reason for the abrupt behavior turn of the tail can be explained by the two obvious facts:

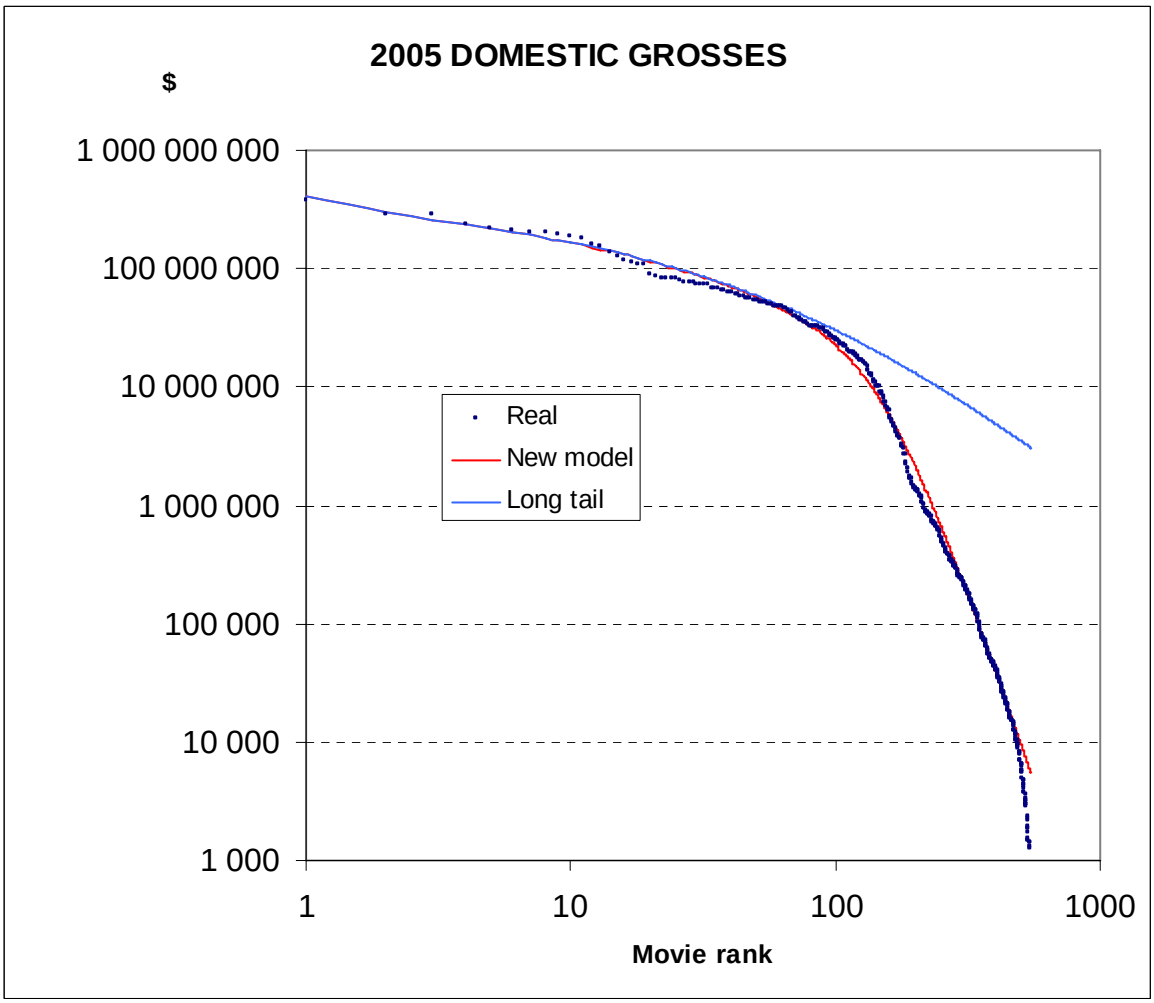
- The number of movies that is able to gain wide publicity at the same time is very limited. Roughly speaking, if a typical (popular) movie gets public attention for 3 weeks and there are 6 movies discussed at the same time by wide audience, the results is about popular 100 movies per year.
- The movie business is much more profitable for the most popular movies than for the less popular movies. One aspect is that the cost per viewer is obviously higher if there are only a couple viewers than if there are, let say, 100 viewers per showing.

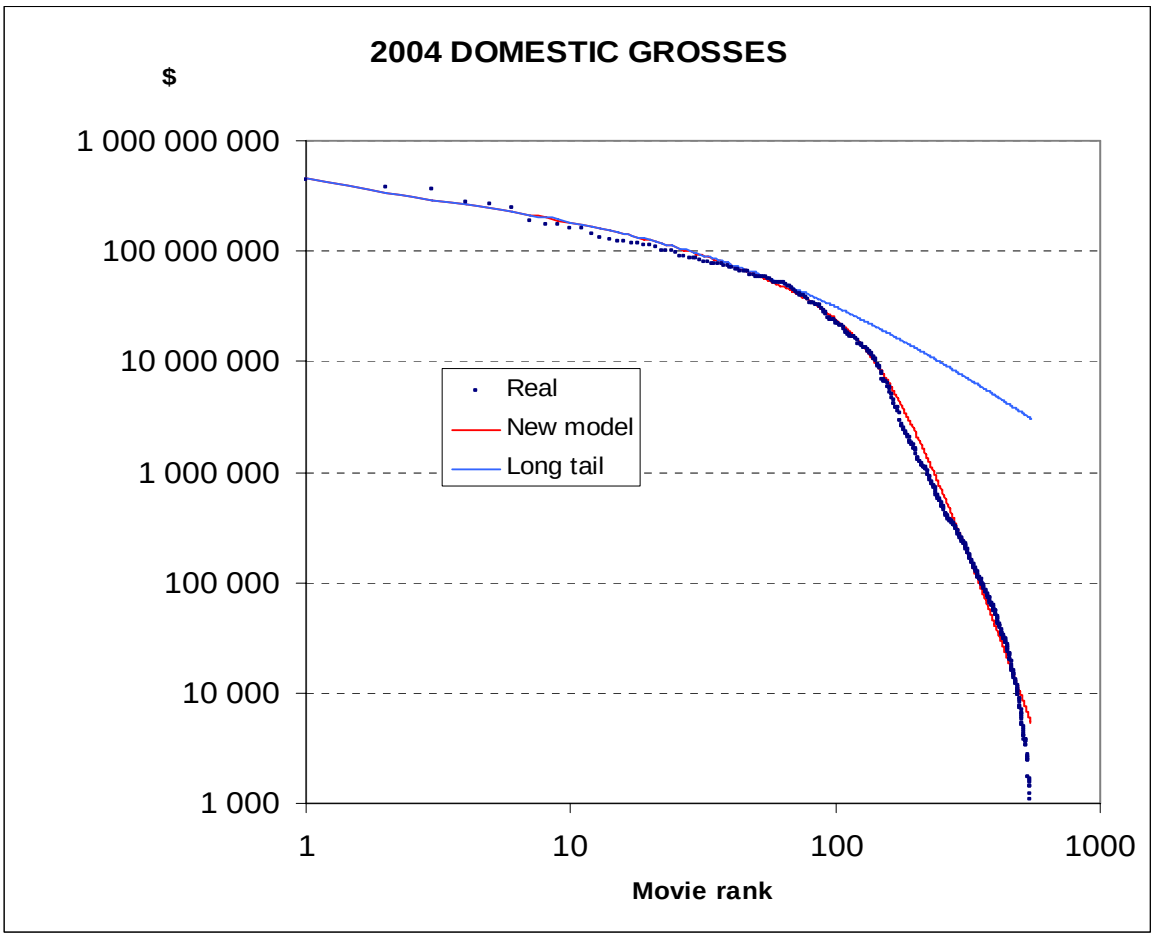
Thus there is a very strong pressure to identify those 100 movies, concentrate the marketing to them and present them on as large screens as possible. Obviously, this process works more efficiently in US than in other countries. As a side effect, those movies that are powerfully marketed in US tend to get the attention in almost all other markets as well. This may partly explain the difference between US and Europe: the 100 most popular US movies will inevitable get attention is Europe as well, but in addition to those movies, there always are local movies that extends the tail of movies.

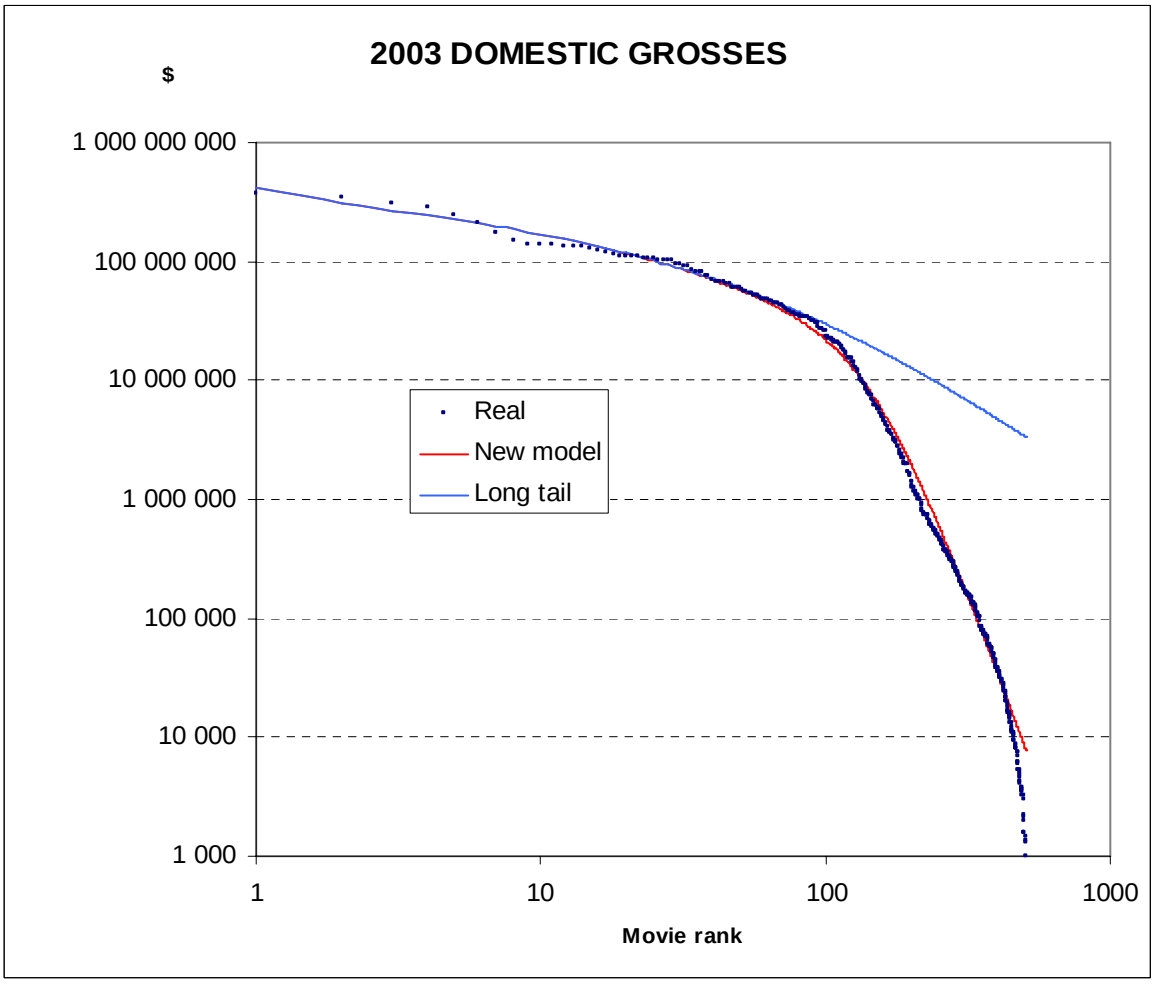
Then there is the question whether some business opportunity is truly lost because of the abrupt turn of the tail. Maybe, maybe not. We may speculate that what is truly lost, is the missed opportunity of serving the specific needs of large minorities. If we consider the true mass market, it might be that the efficient marketing just concentrates the demand on a limited number of movies, but that the process does not necessarily reduce the total demand at all; that is, the demand is just shifted from less popular movies to the more popular movies intentionally selected by someone. This is, of course, the natural viewpoint of large film studios. Still, this process leaves a considerable opportunity for smaller players in niche markets. Parameter $\beta$ may provide a reasonable estimate about the size of the business opportunity.

What the long tail curves (the original and the modified) actually claim is that if the use of the resources (including marketing and allocation movie screens) is distributed more evenly, the long tail of movies would be much longer. That kind of even allocation is, of course, not realistic in the current business model, but may happen if movies were distributed freely through Internet, marketed through peer to peer networks, and watched in home theaters.

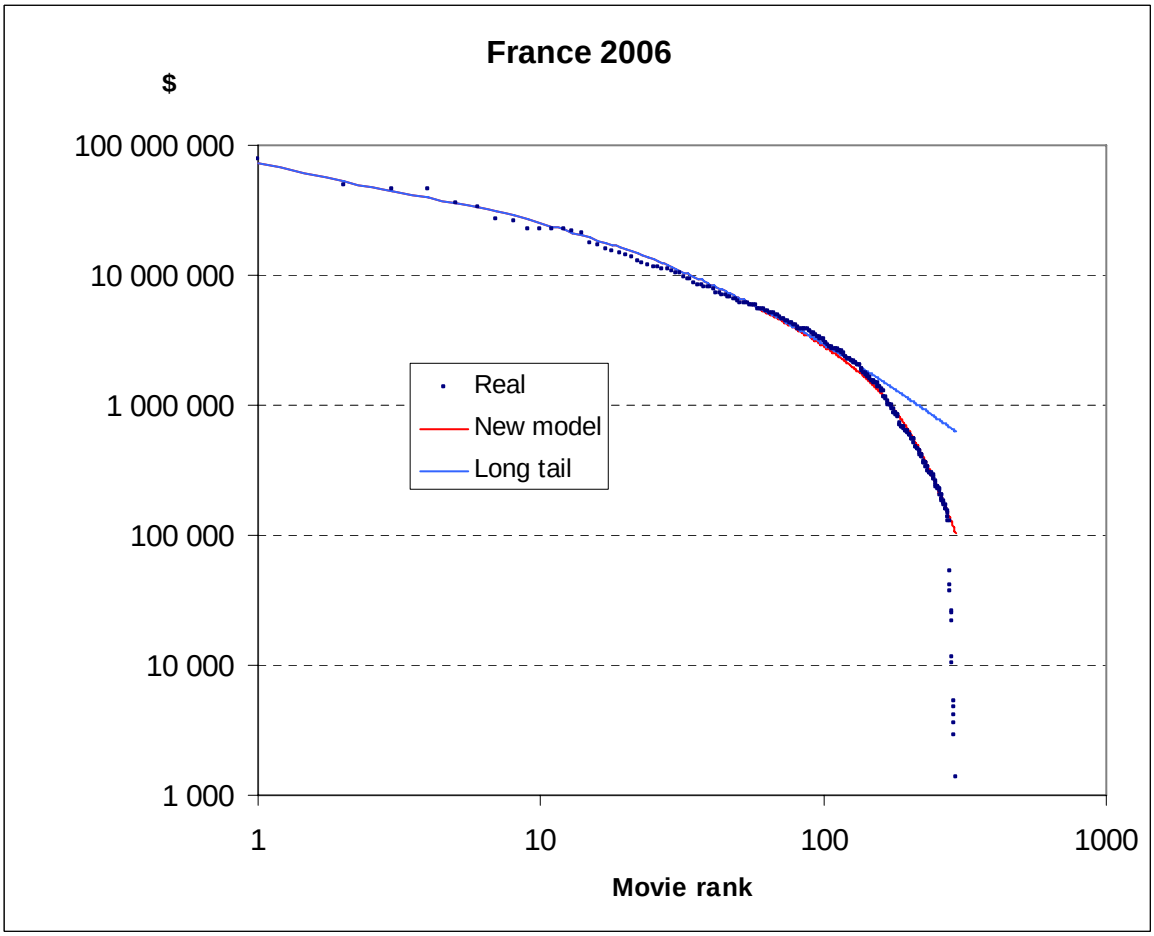

2006 DOMESTIC GROSSES
$
1 000 000 000
100 000 000
10 000 000
1 000 000
100 000
10 000
1 000
Real
New model
Long tail
1
10
100
1000
Movie rank

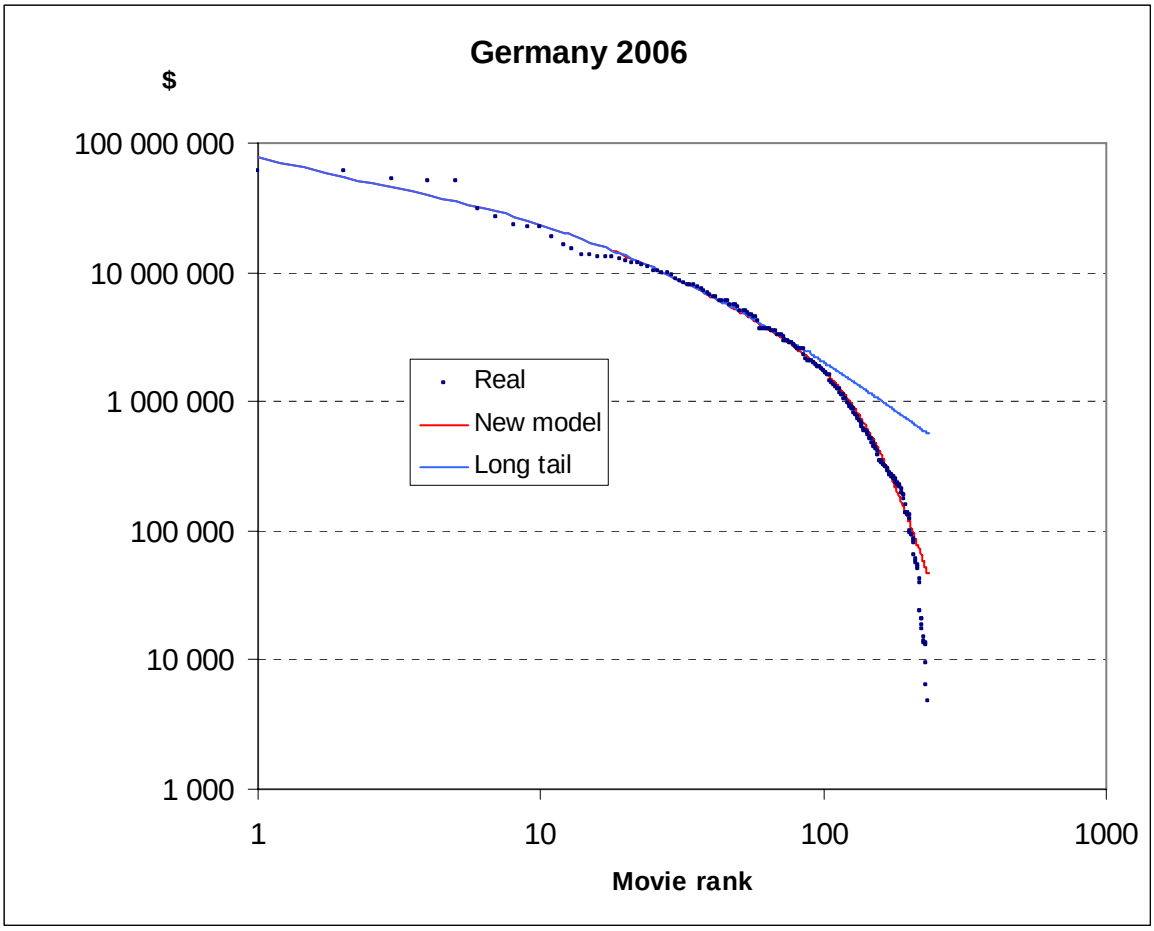

2005 DOMESTIC GROSSES
$
1 000 000 000
100 000 000
10 000 000
1 000 000
100 000
10 000
1 000
Real
New model
Long tail
1
10
100
1000
Movie rank

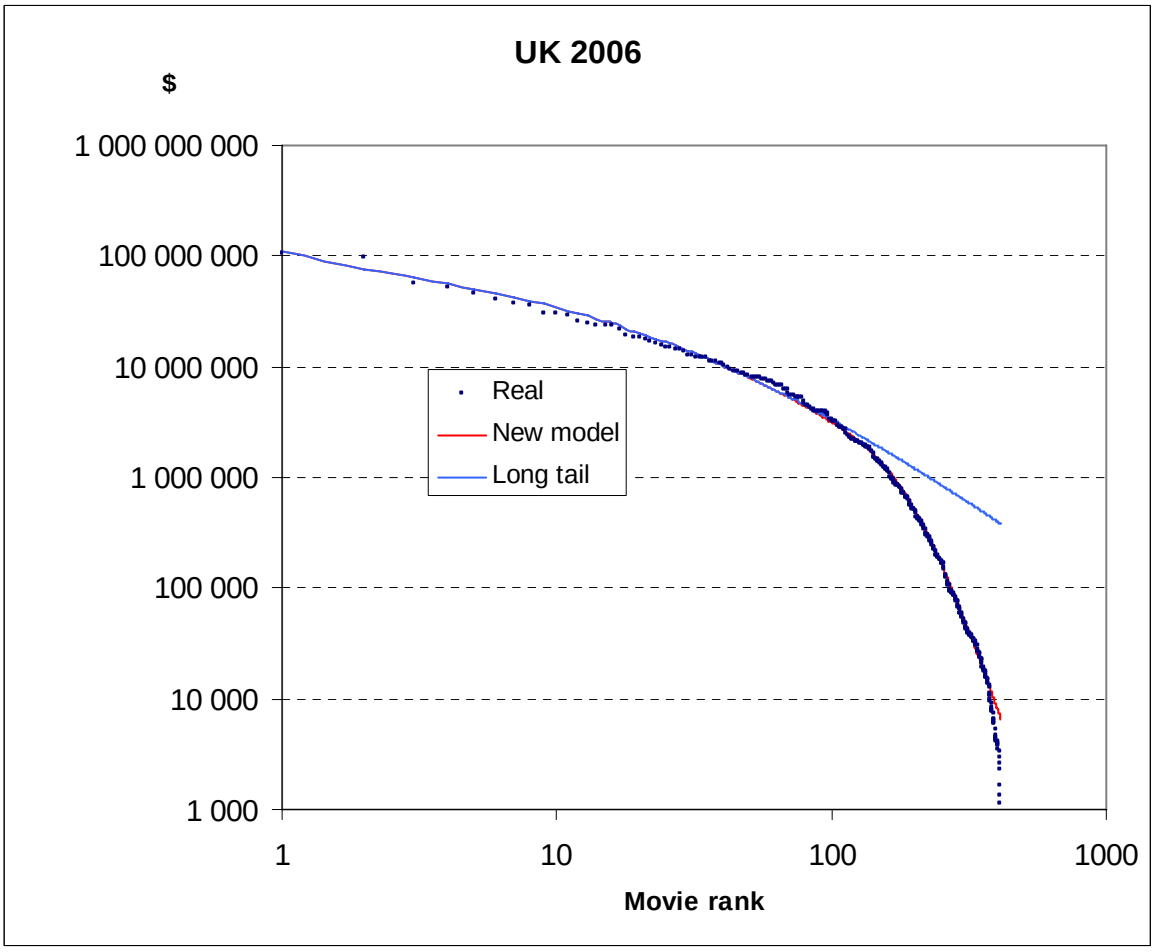

2004 DOMESTIC GROSSES
$
1 000 000 000
100 000 000
10 000 000
1 000 000
100 000
10 000
1 000
Real
New model
Long tail
1
10
100
1000
Movie rank

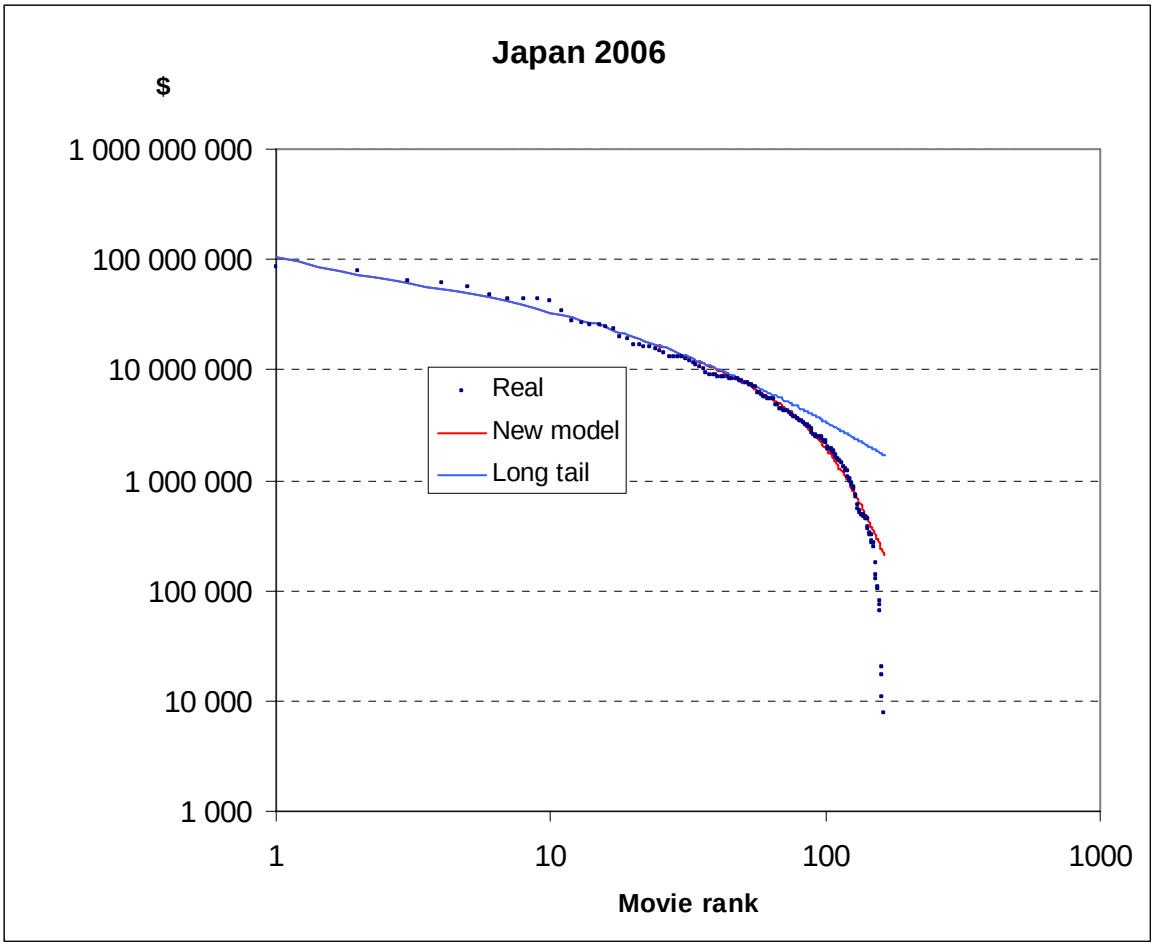

2003 DOMESTIC GROSSES
$
1 000 000 000
100 000 000
10 000 000
1 000 000
100 000
10 000
1 000
Real
New model
Long tail
1
10
100
1000
Movie rank

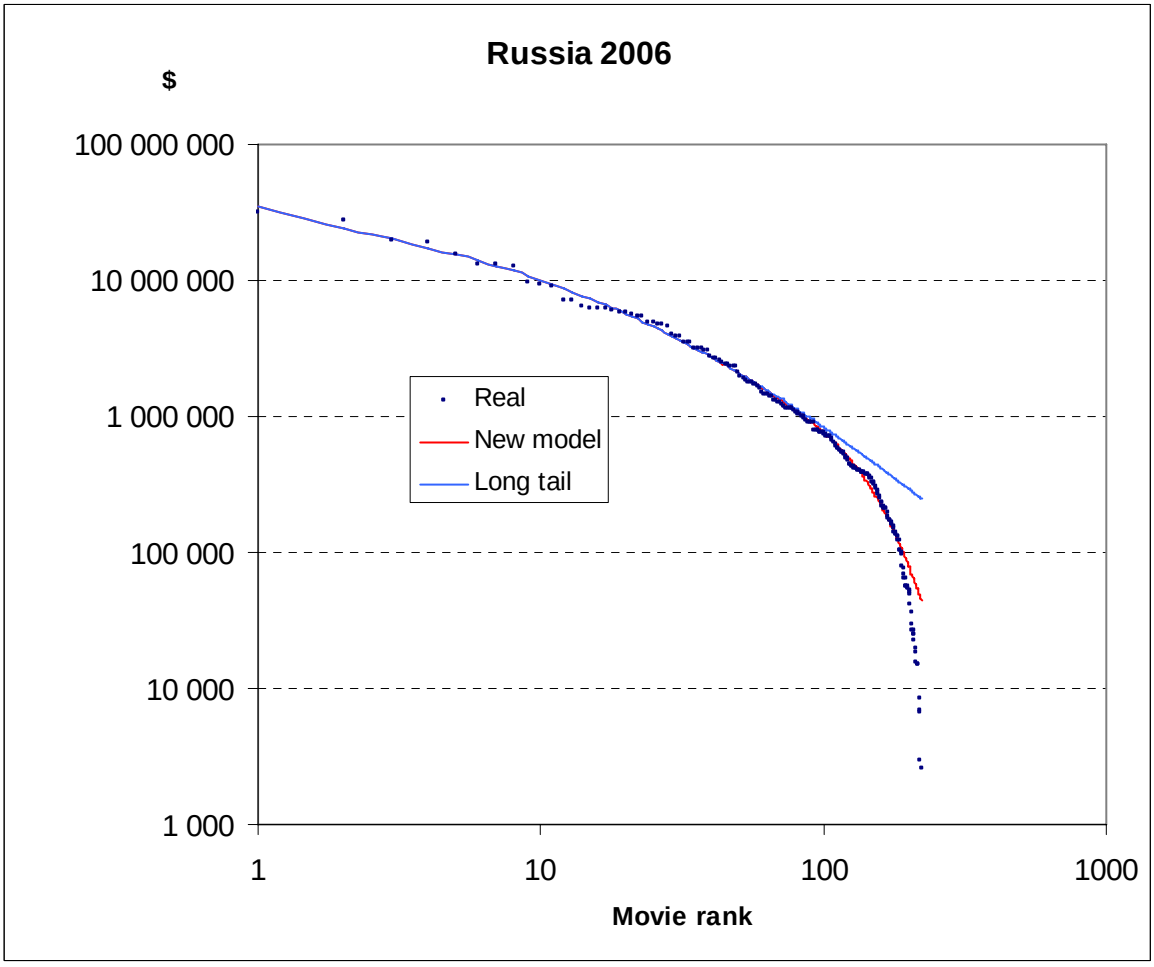

2002 DOMESTIC GROSSES
$
1 000 000 000
100 000 000
10 000 000
1 000 000
100 000
10 000
1 000
Real
New model
Long tail
1
10
Movie rank
100
1000


France 2006
$
100 000 000
10 000 000
1 000 000
100 000
10 000
1 000
Real
New model
Long tail
1
10
100
1000
Movie rank


Germany 2006
$
100 000 000
10 000 000
1 000 000
100 000
10 000
1 000
Real
New model
Long tail
1
10
100
1000
Movie rank


UK 2006
$
1 000 000 000
100 000 000
10 000 000
1 000 000
100 000
10 000
1 000
Real
New model
Long tail
1
10
100
1000
Movie rank


Japan 2006
$
1 000 000 000
100 000 000
10 000 000
1 000 000
100 000
10 000
1 000
1
10
100
1000
Movie rank
Real
New model
Long tail


Russia 2006
$
100 000 000
10 000 000
1 000 000
100 000
10 000
1 000
1
10
100
1000
Movie rank
Real
New model
Long tail