Simulating a Ghost-Writing Startup - Part 2

by Jonathan Chen

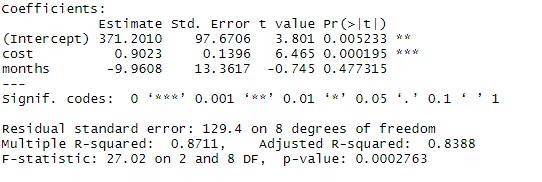
**Abstract:** There is rarely a business endeavor that is profitable from the start. In this follow-up analysis, I will detail using Arena to simulate the interactions between revenue, marketing costs, and the customer fanbase for a book series published on Amazon via ghostwriters. Using the takeaways from my previous paper and newly available marketing and sales data, my task is to determine under what conditions will our company generate a profit. I will first simulate how a standalone book behaves and tune the model to mirror historical data. I will then establish some baseline assumptions to expand the model to handle multiple books in the same series. Ultimately I find that profit is feasible. However, even under relatively optimistic parameters used in the model, it will require a decent overhead investment of both time and money.

**Background/Problem Description:** After a previous analysis where I examined the financial feasibility of our company’s pay structure and business operations, I realized that the nuances of buying a manuscript are overshadowed in importance compared to how a book sells. I also concluded that marketing, whether done consistently or just to renew interest after novelty wears off, was an indispensable investment that must be made despite how bullish I simulated customer traction to be. This meant that in this follow-up analysis I could simplify the model and focus solely on marketing costs and revenue.

I will first construct a simple model with just one book and tune it to mirror the financials of an inaugural book in a series. This step serves as a reality check and will provide insight on profitability in the short term. After analysis I will then generalize to include multiple books and reexamine profitability.

**Main Findings:** One flaw of the previous model was that I underemphasized marketing as a cost component and only focused on book sales as a source of revenue. Revenue comes in the forms of traditional book sales royalty, and royalty from pages read by premium Kindle subscription users. Instead of simulating the royalties of each one separately, one simplification will be simulating the combined sum.

One assumption I must make, was that the currency between marketing and sales is solely based on clicks generated from showing ads. Which is not a horrible assumption considering all our company’s marketing and sales are done digitally. This makes the number of clicks our ads get a function of expenditure and time. It can also be assumed that each click has some averaged out monetary value, or revenue per click (RPC).

To solve for the relationship between ad spend and clicks, I ran a simple linear regression between the two and included a monthly time component to identify any possible trend. The summary is below.

The intercept can be interpreted as the “novelty factor”. Implying that you can get 371 extra clicks for an ad just because it is new.

The cost coefficient signifies that every dollar you spend, we would get .9 more clicks.

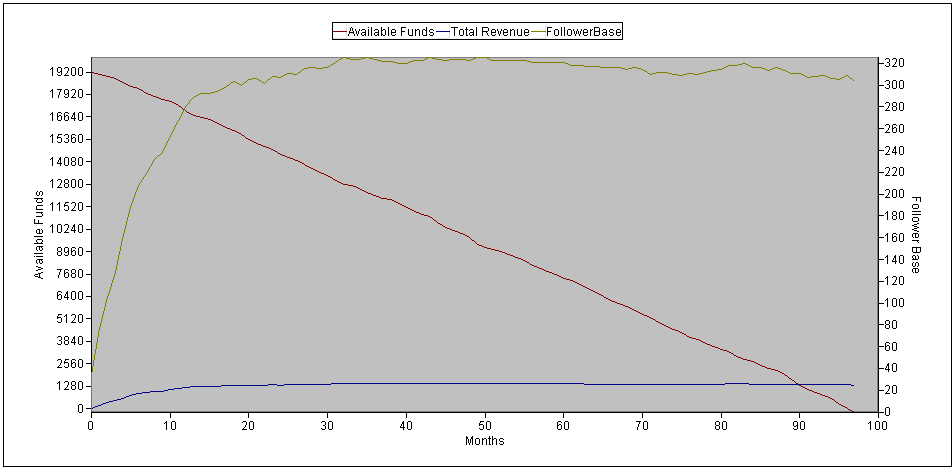
The coefficient for months implies that your potential for clicks decreases by ten every month after initial release.

Even though only cost and intercept are statistically significant to predicting clicks, I will be including the monthly degeneration value as well. It makes sense that over time there will be a natural decline in ad effectiveness.

Now that we have a way to simulate clicks over time, we need to look at RPC. The real data is in blue while the function I used for approximation is in orange.

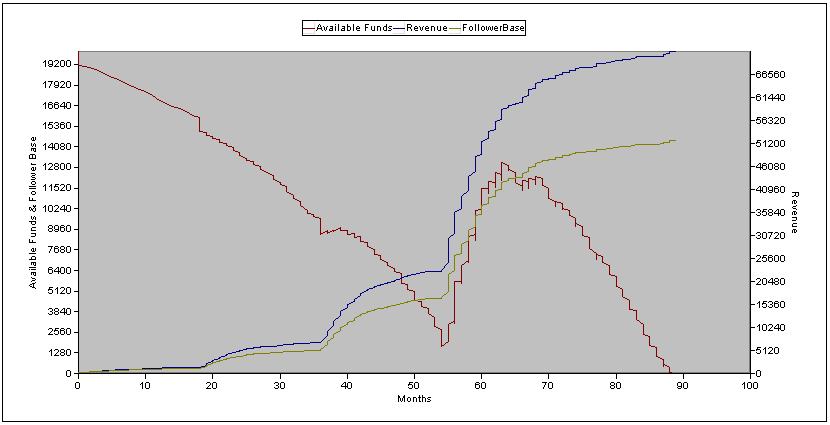
Similarly, to how I handled sales in the previous model, I approximated the decreasing behavior of RPC by using a Poisson distribution. Poisson distributions have interarrival rates that are exponential which take the form of: . Where lambda () becomes my starting sales rate, ‘t’ represents time that has passed, and beta () allows me to control how fast the function decreases. Specifically, I decided to make these parameters: .

I now have all the pieces for an initial run. Like before I will initialize the amount of *Available Funds* to $20k. Having manuscript cost as a random variable did not have much impact on how the last simulation ran, so in this analysis it will be a flat value. Each manuscript will be bought for $2800 of which $700 will be paid upfront and $2100 will be paid as royalty payments overtime to the ghostwriter. To mirror the actual financial numbers, $140 in ad spend will be used for pre-order exposure. Afterwards, ad spend will be a normal random variable with: . *Click Potential* follows the function: . Revenue will be calculated by multiplying *Click Potential* with the current RPC, which will follow the function: Like in the previous analysis, 60% of the royalty will be going towards paying off the remaining $2100 owed to the ghost writer, and 30% will be returned to the *Available Funds*. After the ghost writer has been fully paid off, 72% will go back into *Available Funds*. I will run this simulation until either 100 periods (months) have passed or *Available Funds* runs out. Keeping track of total revenue earned, Available *Funds*, and *Follower Base* (which I will get into) gives us this output.

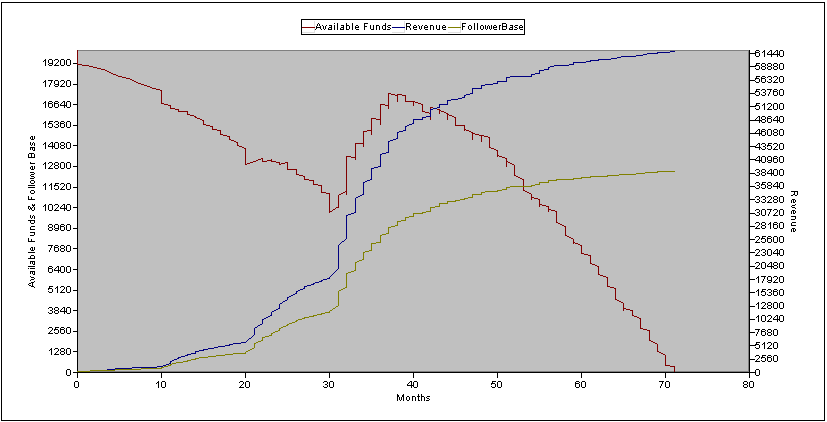
Forecasting over 80 periods from only 12 periods of data will never yield accurate results. But the best we can do to confirm that we at least started on the right foot is by comparing the first 11 periods of this simulation to the 11 periods of actual data that we do have. By period 11, the simulation spent around $2800 to get $1100 in royalties, which is not far off from the actual figure of $3000 that was spent to get around $1000 of royalties. The simulation is more optimistic, but this makes sense as our RPC curve was usually more generous than the actual rate. The output implies that even with consistent advertising, royalty payments for a single book will stagnate around the 12-period mark. All the while revenue continues to drop every period as we continue investing in advertisement. So far, not so good.

But advertisements do not just generate income, it also generates a follower base. Not every consumer will become a follower, but those that do will much more likely buy subsequent books in the series. Factoring in the lifetime value of a customer, the road for financial stability becomes clearer. Investing in and creating a consumer base for the first books in a series, will eventually lead to more sales later in the series. And the more books there are in a series, the greater the lifetime value of each customer.

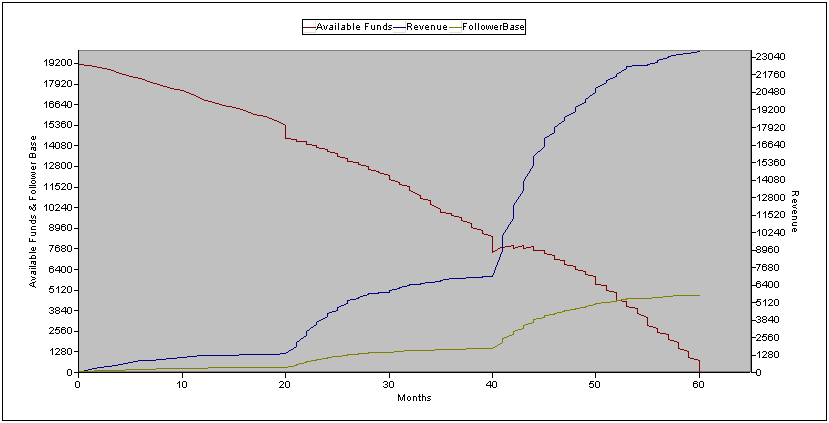
I took a simplified approach to working pre-existing followers into book sequels. The number of followers gained is proportional to the revenue from the same period. Every time a book is published, I would increase the click potential by the number of current followers. This does not change anything for the first book, but every subsequent book will start off with more click potential. Like in the previous analysis, I also decided to reset the counter for each book every time a new book in the series is published. Intuitively, this is to mirror associated products (be it books, movies, or etc.) getting attention again when its sequel is released (think movie marathons, collectors bundles, etc.). Finally, according to this [article](https://kindlepreneur.com/calculate-series-read-through/?fbclid=IwAR36QTuJPgtuflDQQkLmrZT1DpGyEaKd6-bDJwTSt9hi7PpHRGadMhkFGvo) that the company’s founder shared with me, most authors can expect around a 60% read through rate from the first book to the second. Which makes marketing a series of books more efficient than if they were all standalone. For this simulation I will use a 60% read through rate for all books. Assuming there are four books, selling one book also means selling 60% (.61) of a second, 36% (.62) of a third, and 21.6% (.63) of a fourth book.

Rerunning the model after those adjustments gives us the following new output. This time I will track *Available Funds* and *Follower Base* on the left axis due to similar scale between the two and *Revenue* on the right. As before, I will be simulating until either 100 periods have passed, or until *Available Funds* runs out. I will also be operating under the assumption that there are four books to this series, with each subsequent book taking 18 periods to write.

The key takeaway from this output is that once a customer base has been successfully established, a book series will become profitable. We see this inflection point around period 54. Which is coincides to the release of the fourth and last book in the simulated series. We then see profits for around ten periods, before the ultimately the monthly decrease stagnates royalties again and drives *Available Funds* down to zero. This happens because we have no more books to release at this point.

It is interesting to note that this model implies that timing between book releases is important. Shorter interarrival times between books lead to quicker profitability, but this does not always translate to more revenue and financial health. It may be tempting to shorten that 54 month period of losses, but the benefit may not always outweigh the cost.

For example, keeping everything else *ceteris paribus* and changing only the time between publications, if we decreased the waiting period of 18 to a waiting period of 10 we would instead get this output.

With this configuration, we see profits earlier around period 30 compared to period 54. However, we also run out of *Available Funds* quicker (70 vs. 88). Additionally, both *Follower Base* and *Total Revenue* are lower, even when compared to period 70 of the previous model (12.5k and $62k vs. 13k and $65k).

On the other hand, waiting too long also has negative effects as well. Increasing the waiting period to 20 (only 2 more months than 18) gives us this output.

This configuration does not ever become profitable and terminates at period 60. Moreover, the values at termination are no where near the previous two’s.

There seems to be a sweet spot in releasing books. Releasing them too close to each other cuts the attention lifespan of each individual book and not all royalty sales are realized. This may be akin to market cannibalization. Releasing the books too far from each other runs the risk of falling out of relevancy and going bankrupt.

The number of books in a series comparable has less impact on the overall behavior of the model. In general, the more books you plan to publish, the longer the total lifespan of the series. Though hitting bankruptcy prematurely overrides the behavior of the model regardless of how many books were planned.

There last thing of note is that the time to pay off the $2100 owed to the ghostwriter takes upwards of 54 months, even in the best model. Which is not a realistic turnaround rate if we do not want angry ghostwriters complaining about not getting compensated. The only way for this number to decrease is if we both drastically decrease the time between books, and drastically increase the number of books in the series. This however sparks questions of feasibility and quality. And with no prior information on what the behavior of quick turnaround books would look like, it is unlikely that the current model would be accurate in dealing with such a corner case.

**Conclusions:** After making and running this simulation, it makes sense why many business endeavors take time before becoming profitable. Profit often comes with mass and scale. Scale only comes when the product has had enough exposure. Lastly, exposure only comes with time and investment. Anecdotally, this makes sense. I cannot remember getting into any book series in its infancy. When I started reading Harry Potter, the third book had already been released. I’ve also had similar experiences with Artemis Fowl and even the Hunger Games. Another interesting result was that there was a sweet spot when it came to releasing books (at realistic quantities). It surprised me that I had unwittingly simulated a case of market cannibalization. For next steps, one would be to do more research in the interaction between follower base and clicks. Currently I believe that interaction is too naïve. Furthermore, once sales data from the baseline book’s sequel comes back, I am sure I will have to retune the model.