Simulating a Ghost-Writing Startup

by Jonathan Chen

Abstract: There are many variables that inform how a company operates wholistically. In this paper, I will use Arena to simulate the business operations of publishing books on Amazon via ghostwriters. The company uses a novel payment structure that rebalances a ghost-writer's payment potential. Using previously established statistics, my task is to simulate and determine the efficacy of this specific business model. In the rest of this report, I will first set up a naïve simulation with conversative values for the variables and then make more nuanced changes. I found that this payment structure rebalances payment potential so much in the company's favor, that it may limit future business opportunities. Specifically, the ghost-writers are not renumerated as quickly or consistently as they might be if paid traditionally.

Background/Problem Description: Traditionally, paychecks would be issued bi-weekly as the book is being written, though the simulation in this project will handle the paycheck as a one-time payment prior to writing. Doing this circumvents the need to budget dynamically, which would be used to avoid over-expenditure and the inability to pay writers for their work. However, our company has an alternative payment method that needs to be tried and evaluated. Our company does not pay the ghost-writers' paychecks in full, instead making up the remainder through royalties, and allowing ghost-writers to earn a permanent royalty later on. This is not something most companies offer and is a selling point for this model. The tradeoff for the writer becomes making less than the agreed upon rate in the short term, for a chance at making much more in the long run if the book sells enough.

To this end, after considering given financial variables and resources, I will construct a naïve model of financial performance of the company. Which after analysis and peer evaluation, I will refine for more realistic sales behavior.

Main Findings: Because the focus of this simulation is primarily the financial viability of the current business model, we will look at the various ways business operations impact our *Funds Available*. Our first analytical priority is to determine sources of *cost* and *revenue*.

Disregarding marketing, the main cost of doing business is "buying the books" that we intend to publish. As a ghost-writing company, the crux of the business revolves around paying writers to finish desired manuscripts for books. This will be referred to as, "buying the book" for the remainder of the report. The associated costs can be determined by the following variables, which were kept simple for the first round of simulation.

Variable	Value	Justification
Word Count Target	50,000	A lowball wordcount estimate given by the founder, for the types of books he would like to publish.
Words Written Per Hour	375	A standard writing speed estimate given by the founder.
Hourly Pay Rate	UNIF(15,25)	$15/hr being the potential minimum wage and $25/hr being a suggested cap suggested by the founder.

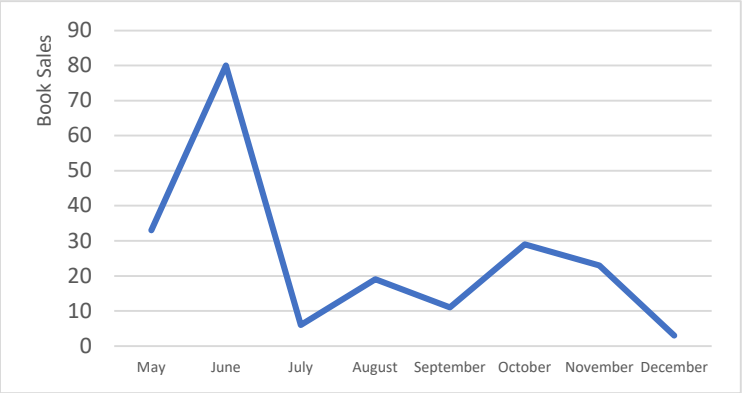
These three variables give us the cost of producing the manuscript of each book. Dividing *Word Count Target* by *Words Written Per Hour* tells the labor hours. Multiplying that number by an hourly rate gives us the total price we are buying the book for.

While the sum of the paychecks would *normally* equal the entire cost of the writing work, once the total cost is determined for the book, only a fraction will be paid as a paycheck. The rest of the cost will be paid off through royalty sales of the book once it has been published. While the benefits for the writer are a little dubious, the benefits for the company are more operationally obvious. By requiring less of an upfront cost to buy a book, this frees up cash to be used for other investments.

Quantifying revenue on the other hand is little less straightforward. According to nearly every Business 101 class: *Revenue* = *Price* x *Quantity*. And the revenue numbers used in this simulation were as follows:

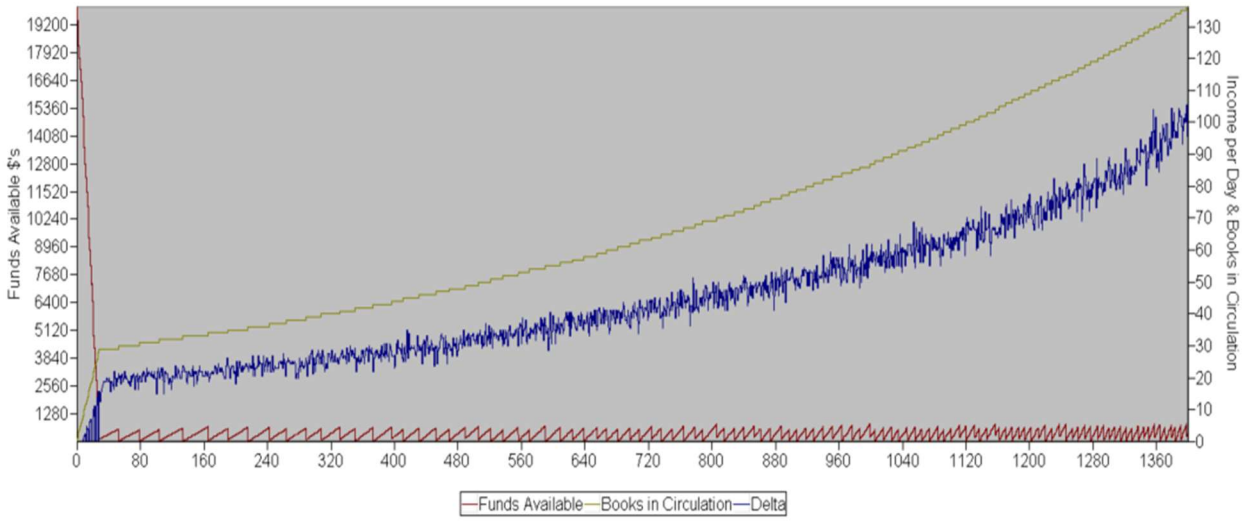
Variable	**Value**	**Explanation**
Listing Price	$3.99: Baseline Estimate	According to the founder, this is a rather standard rate for books around 50k words on Amazon.
Royalty	$2.63 = Listing Price * 66%	After Amazon takes its commission, this is an estimate of how much the company receives per book. But not all of it will go back into our *Funds Available*.
Paid to Writer (Pre-Breakeven)	$1.58 = Royalty * 60%	The writer will be getting 60% of the Royalty as complementary payment, until cost is fulfilled.
Paid to Company (Pre-Breakeven)	$0.79 = Royalty * 30%	While the writer is being paid, the company's *Funds Available* will only receive 30% of the Royalty.
Paid to Company (Post-Breakeven)	$1.89 = Royalty * 70%	After the writer is finished being paid, the company's *Funds Available* will receive 70% of the Royalty.

Quantity is harder to estimate, as it usually comes in the form of demand and I am not very familiar with the economics of book sales. Instead, I created my assumptions and baselines from the time series sales data from a previously published book.



Despite the Poisson-esque distribution of the sales data over time, for my first run-through I decided to keep it simple and use an average rate as a stand-in for demand. There were around 244 days between the start of May and the end of December 2020. Within this period, 204 books were sold. 204/244 comes to roughly 0.83 books sold per day. And *voilà,* instant quantity substitute.

Now that we have revenue and cost, we can finally start simulating. From an operational point of view, the book and its writer can be treated as a single entity. It is assumed that there is no shortage of writers willing to write books and that it is in the company's interest to buy books whenever possible. Once a book has been written there will be an unpaid cost associated with it, which will be decremented every time it sells by a portion of the royalty. Every book has an 83% chance to sell every day, to align with our average. Running the simulation for 1400 days with an initial investment of $20k gives us the following behavior for *Funds Available*, number of books purchased, and cash inflow.

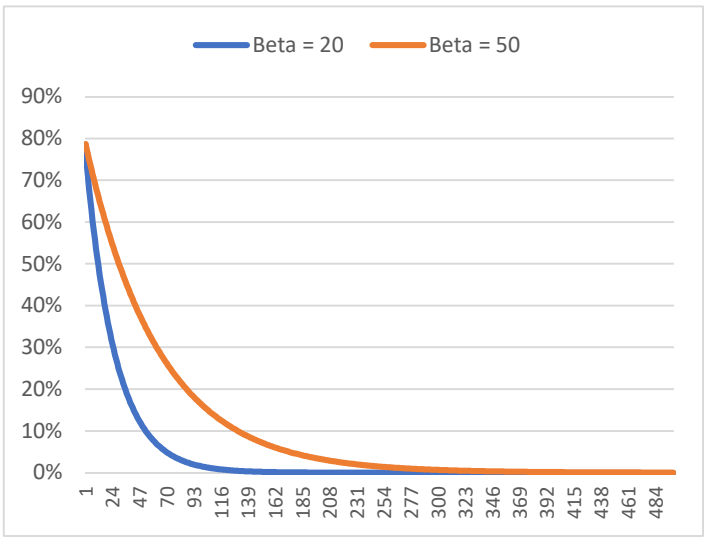


These results are not too surprising. *Funds Available* spikes down whenever we can afford a new book, which increases in frequency over time. Our initial investment is invested completely into new books, which causes a huge downwards spike in funds and a huge increase in books circulating. After this point, both cash inflow (Delta) and books in circulation increase at an increasing rate over time.

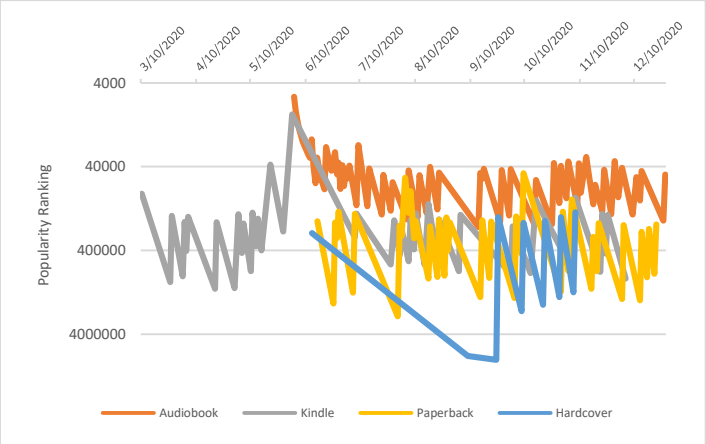
This naïve simulation implies that the current business model works well under the established assumptions and turns a profit. But how much of this success can be attributed to the unrealistic expectation of a constant selling rate of 83%? And how much of this exponential growth can be attributed to our amazing continual ability to recruit writers who are willing to wait an average of 1280 days (nearly 3.5 years) for their full payment to be actualized? This second realization is worrying. Not only because 3.5 years is a very long time for anyone to wait for payment, but also because our constant rate of 83% is already an optimistic assumption.

For my next model, I need a more accurate way of simulating demand. After conferring with the founder, I also decided to factor in popularity and the potentially sales compounding effect of having Amazon recommend other books published by the same source.

Modelling demand is mathematically straightforward. The original sales graph looks like it has a Poisson distribution, and it makes intuitive sense that book sales are independent from each other. Poisson distributions have interarrival rates that are exponential which take the form of: $\lambda e^{-\lambda t/\beta}$. Where lambda ($\lambda$) becomes my starting sales rate, 't' represents time that has passed, and beta (β) allows me to control how fast the function decreases. For example: setting $\lambda = .8$ & $\beta = 20$ gives the blue line and $\beta = 50$ gives flatter orange line. Both of which result in the decreasing rate desired.

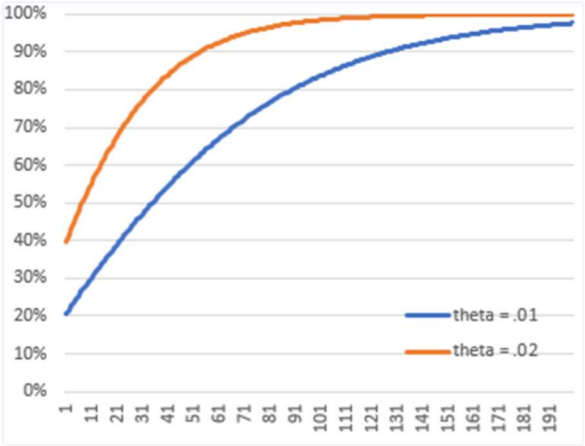


Factoring in popularity calls for a bit more speculation. According to this article, search ranking on Amazon directly impacts product visitation. This makes sense, the more popular your book is, the more clicks you will receive, and more clicks translates to more potential sales. Furthermore, the more clicks and sales a book gets, the more it'll be recommended to customers, increasing its popularity. To get a better mental grasp on popularity, I looked at the previously published book's popularity ranking over time.



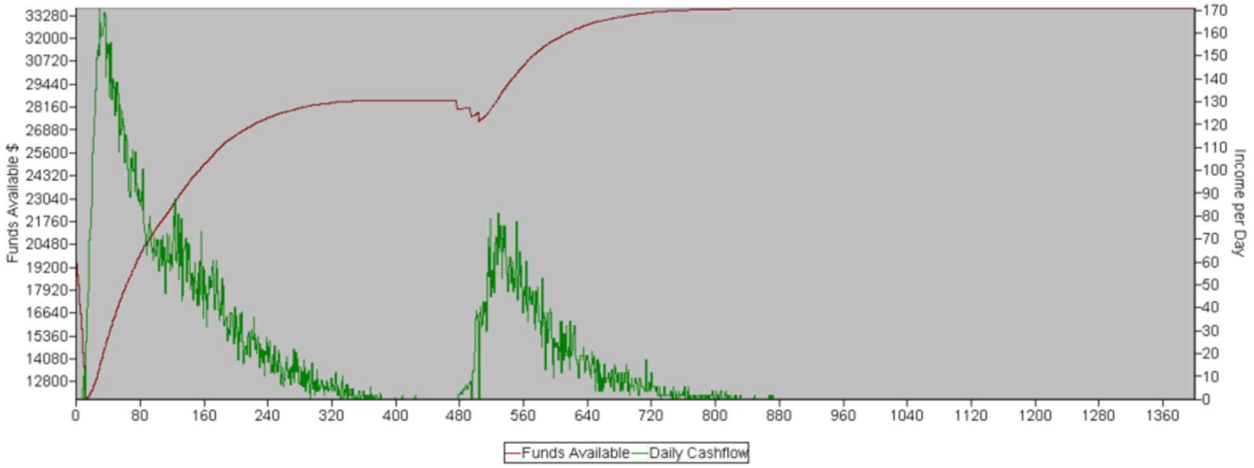
The main takeaway from this graph is that popularity increases and decreases on a logarithmic scale, and is at least somewhat related to sales. This scale implies that it is easy to climb the charts in the lower rankings, but harder to achieve higher rankings. The popularities peak in June, which was also the peak in sales and pages read. It can also be assumed that popularity decreases over time if a book does not sell well.

If we are to believe the article's assertion, *"if you get your book to rank #1…then you can expect about 540 clicks per month"* then achieving a top rank equates to 540 clicks per month, or roughly 18 clicks a day. Multiple clicks a day can be interpreted as multiple sales opportunities a day. But to match how hard it is to reach rank 1 in popularity, I must also pick a function to make it hard to reach that 18 times multiplier. For this, I chose the hyperbolic-tangent function $\tanh \alpha\theta$ $\alpha =$ popularity ranking, and $\theta =$ publisher popularity. Increasing α causes the function to approach 1, and as θ increases, it becomes easier to achieve multiple clicks per day. The output of the function represents a percentage of the potential eighteen chances per day.



A separate article shared by the founder implies that having books that are part of a set helps sales compound on each other. Loosely generalizing, having more books out at the same time in the Amazon bookstore should increase clicks because Amazon will recommend your other books.

Putting that all together: popularity affects how many sales opportunities there are a day, sales rate decreases with time, popularity increases with sales and decreases in absence of sales, and having more books in circulation increases sales opportunities. For this run, instead of buying every book possible this time, I decided instead to put only 13 into circulation, which is realistically enough to fill the entire "related products" ribbon on an Amazon page. After running for 1400 days we get the following graph.



The first thing that should stand out is the decreasing daily cashflow. This makes sense as over time, regardless of how popular a book has become, the sales potential will follow its exponential function and go to zero. We can interpret this as general attention for a book having a shelf-life, which can be attributed to real life passage of time and competition for the spotlight. The decreasing cashflow corresponds nicely to the plateauing total funds.

The next thing to note is the structural break around time 480. Initially I realized that the exponential function will inevitably decrease sales potential of each book to zero. I then decided to include a module simulating some sort of marketing intervention, which resets the time counter for the book at the cost of $500 per instance. Think of this as reigniting marketplace attention for the book. You can see this expense when the *Funds Available* dip, around the same time cashflow increases again. Marketing was only performed if a book failed to fully pay off prior to losing complete public interest. The last thing to note, is that by simulating multiple sales opportunities each day which scaled based on popularity and past sales performance, each book finished paying off way faster than before. The new average being 220 days. While still not an ideal waiting time to be fully paid off, it is markedly better than the previous estimate.

Conclusions: An unexpected takeaway for me, is that marketing is a necessity to compete in today's marketplace. Another is that another look into our pay structure might be warranted, if we don't want potentially disgruntled writers advertising their displeasure to the rest of the writing community. There are still many improvements that can be made to this model. I could flesh out the effect of publisher popularity. I could do more research on the effects and costs of advertising. I could collect more data to get a better grasp on the full effect of popularity on a book's sales. Despite the last simulation's optimistic results, I worry that the positive effect of popularity was exaggerated. Lastly, I could work in the effects of bundling books together as sets, which should save on marketing costs and increase sales of the type.