The Binding of Isaac: Rebirth - Identifying Underlying Covariates Propelling Humanity to Engage with the Cult

1 Introduction

1.1 Background and Motivation

Understanding and forecasting demand is vital to the operations of any firm, and it is especially vital to understand the significant underlying factors that may affect demand in the future. Using adoption data from the first 111 weeks after the launch of *The Binding of Isaac: Rebirth* on Steam, I fit a variety of models and covariates on the calibrating data (weeks 1-61) and compare their fits across both the holdout period (weeks 62-113) and the entire dataset. After fitting using the Exponential, Exponential Gamma, Weibull Gamma, Shifted Weibull Gamma, and the Shifted Weibull Gamma with covariates on Sales, Christmas, Sales of the accompanying *Afterbirth* expansion, and Patches, I find that the Weibull Gamma with the covariates of Sales, Sales of *Afterbirth,* and Patches to provide the best fit.

N	1000000
r	1102.951
alpha	763587.4
c	0.391669
B_sales	0.958402
B_Afterbirth	0.751768
B_Patch	0.510568
LL Sum	-87211.4
ISM	0.399622
OSM	0.677342
TotalM	0.547084
BIC	174505.6

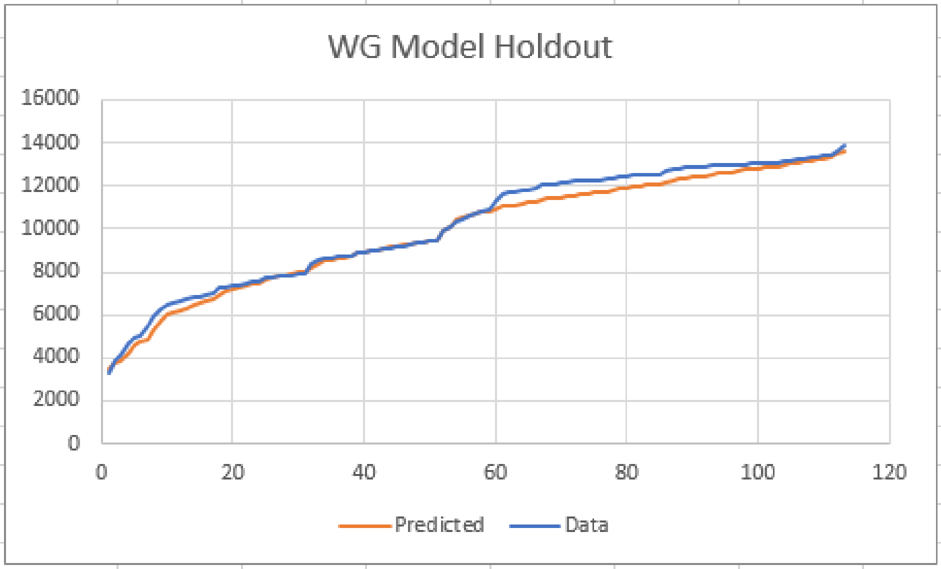


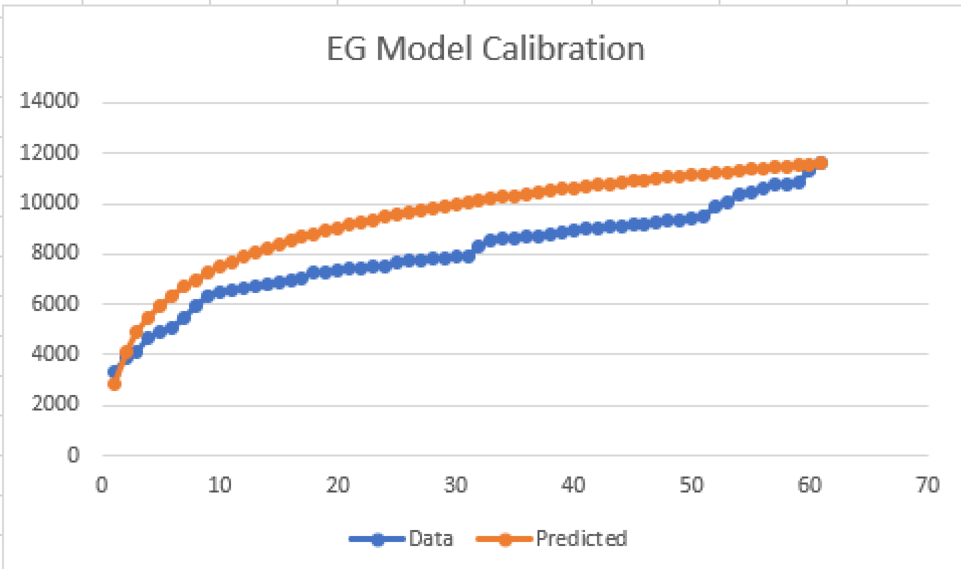
Figure 1: The final Shifted Weibull Gamma model containing covariates of Sales, Sales of *Afterbirth*, and Patches

1.2 Data and Methodology

This paper analyzes the aforementioned adoption data on this video game to demonstrate the capabilities of the models I fit and the methods to obtain better fits and better forecasts. Specifically, using the raw data consisting of only the date of adoption and the number of adopters, I constructed many models, listing only the relevant ones above. I first fit the data to Exponential Model as a baseline model to compare results and construct further models from before moving onto more sophisticated models. I moved onto the Exponential Gamma to provide a mixing distribution and model the heterogeneity in the data set. Understanding the importance of duration dependence, I incorporated the Weibull Gamma to both preserve heterogeneity and allow my model to account for the effect of time on adoption. To account for the pre-order period, I utilized a shifted version of the Weibull Gamma, to provide an estimation of adoption before our recorded data point and allow for a truer granular week-by-week view of the adoption. The next relevant models I fit were the Shifted Weibull Gamma with different covariates, attempting to determine if there were any other extraneous data I could account for to craft a better model.

II Modeling Adoptions of *The Binding of Isaac: Rebirth*

We begin by fitting the Exponential model to the calibration data (weeks 1-61) of adoptions. However, the model fits quite poorly without allowing for heterogeneity, and we will move onto the Exponential Gamma. The following lined scatterplot and table summarize the resulting model. We can see how the actual recorded observations under "Data" compare against our "Predicted".



N	1000000
r	0.002334
alpha	0.41686
LL Sum	-104569
ISM	139.3112
OSM	476.1028
TotalM	294.295
BIC	209165.7

Figure 2: Observed versus expected (under the Exponential Gamma model) number of adoptions over the calibration period

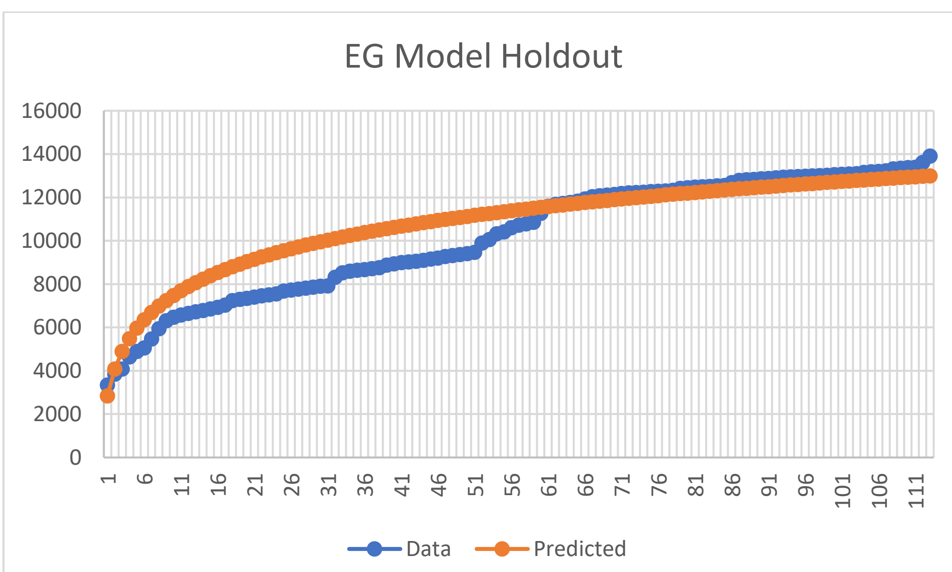
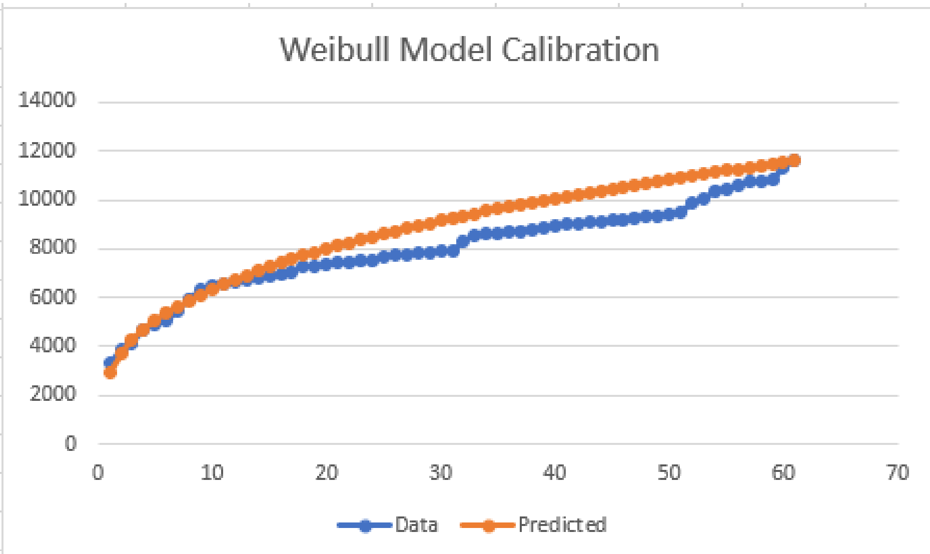


Figure 3: Observed versus expected (under the Exponential model) number of adoptions over the entire period

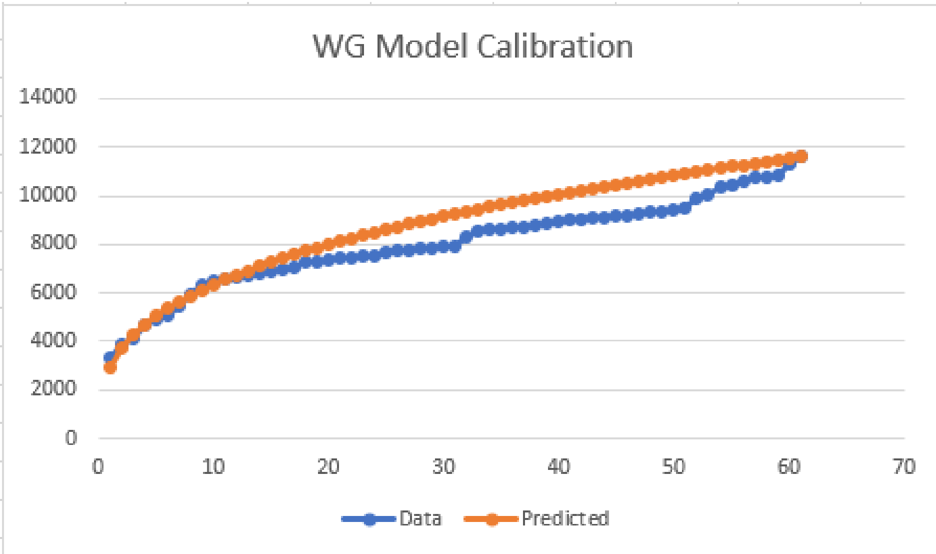
We see that with a low r, there is a fair amount of hetereogeneity and looking at the different MAPEs, we see that the model fits the in sample data the best and fits quite poorly with the out of sample data. It is clear that we should find a better model to fit the data.

The next model we attempt is the Weibull and Weibull Gamma. The Weibull allows for duration dependence and the Weibull Gamma also allows for hetereogeneity. The following lined scatterplot and table summarize the resulting model.



N	1000000
lambda	0.00294
c	0.33516
LL Sum	-103483
ISM	129.6385
OSM	502.6805
TotalM	301.3039
BIC	206993.5

Figure 4: Observed versus expected (under the Weibull model) number of adoptions over the callibration period



N	1000000
r	106.7096
alpha	36300.06
c	0.335173
LL Sum	-103483
ISM	129.6382
OSM	502.677
TotalM	301.3021
BIC	207007.3

Figure 5: Observed versus expected (under the Weibull Gamma model) number of adoptions over the callibration period

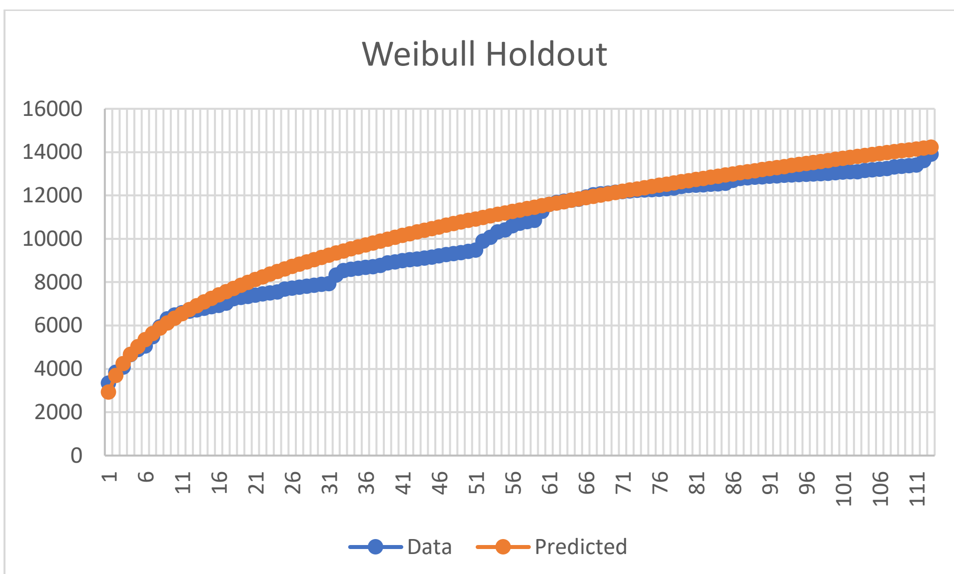


Figure 6: Observed versus expected (under the Weibull model) number of adoptions over the entire period

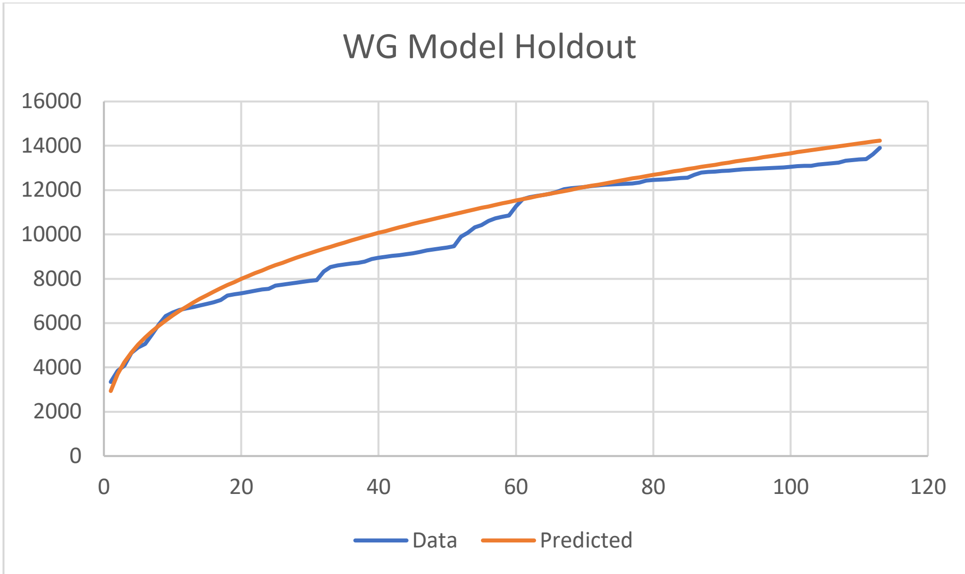
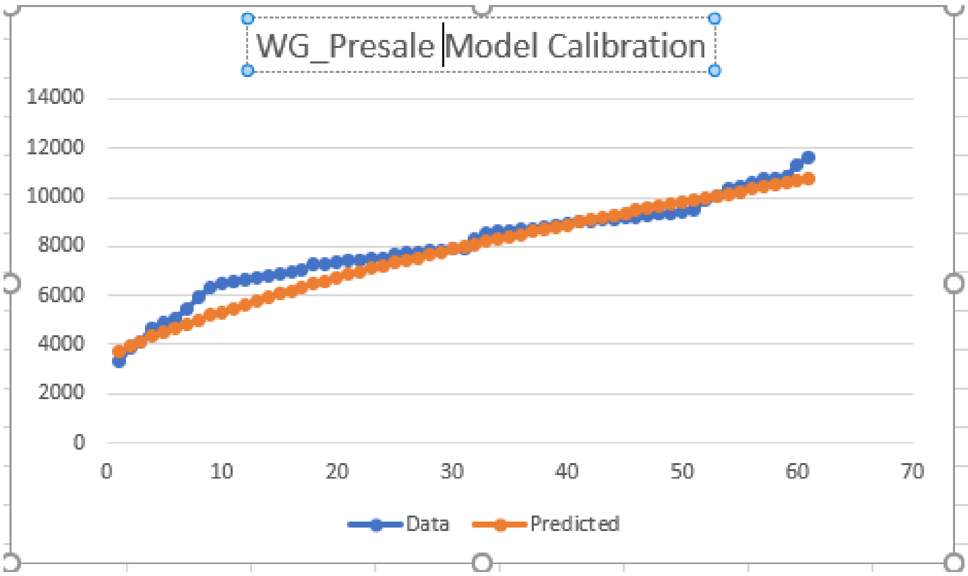


Figure 7: Observed versus expected (under the Weibull Gamma model) number of adoptions over the entire period

These models have a much better fit than the mere exponential gamma, with a much lower BIC. However, we also see that these two models are nearly the same, and that is due to the heterogeneity, or lack thereof. This causes the Weibull Gamma to collapse into the Weibull, as the only difference between the two is the addition of the mixing distribution to allow for incorporating heterogeneity. If there is no heterogeneity, then there will be no difference between the two models. After incorporating duration dependence, we see that there is a high r, indicating high homogeneity. It seems as if all the heterogeneity we observed in the Exponential Gamma was due to the effects of duration dependence, not the propensities within the individuals themselves in the population. However, we also need to account for the behavior of the population based on the pre-order period prior to the release date.

We use the Shifted Weibull Gamma to account for the effects of the pre-order period. Using this model allows us to estimate the "adoption" of individuals in the pre-order period, allowing us to obtain a more granular view of all the adoptions in the data recorded in the first period. The following lined scatterplot and table summarize the resulting model.



N	1000000
r	3469.641
alpha	2925281
c	0.521513
LL Sum	-88539.6
ISM	0.939936
OSM	1.593847
TotalM	1.287145
BIC	177120.7

Figure 8: Observed versus expected (under the Shifted Weibull Gamma model) number of adoptions over the callibration period

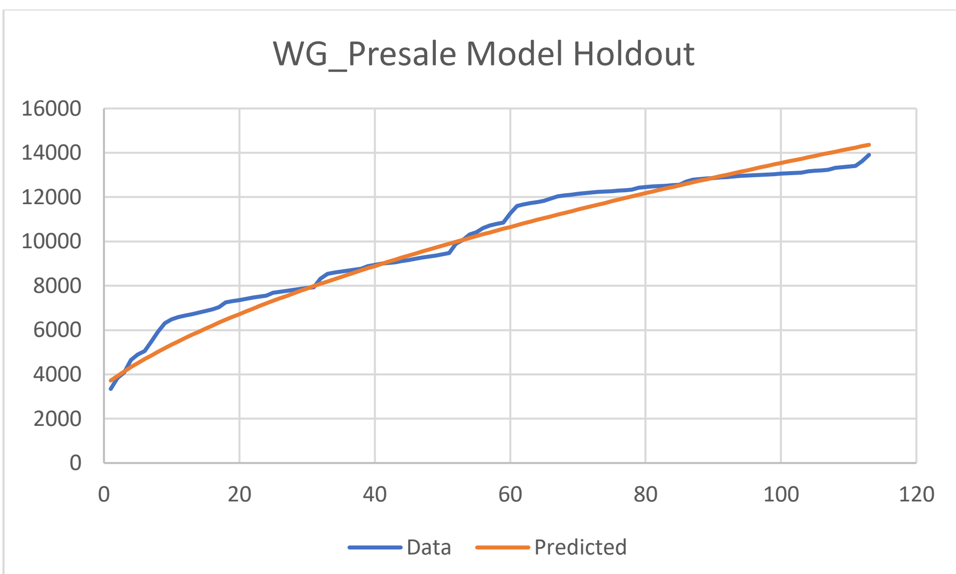
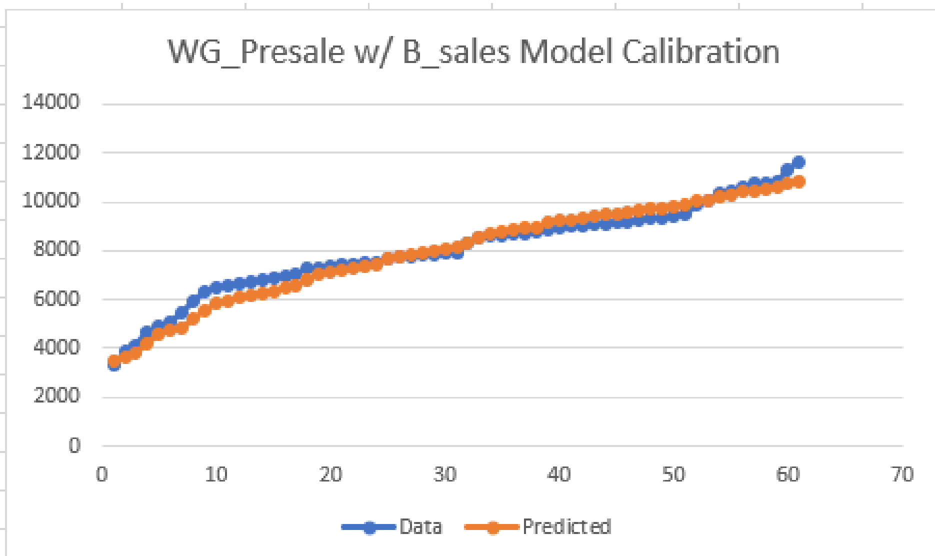


Figure 9: Observed versus expected (under the Shifted Weibull Gamma model) number of adoptions over the entire period

Comparing the BICs of the Weibull Gamma and Shifted Weibull Gamma, we see that the Shifted Weibull Gamma is a much better fit. We see homogeneity increasing with a higher r and the effects of duration dependence increasing as we add in those extra weeks. This is due to the addition of further weeks to explain and model the differences in the data, emphasizing the explanatory effect of duration dependence and drawing farther away from heterogeneity as an explanation. However, the out-of-sample fit is still significantly worse than the in-sample fit and we can visually see deviations in the data that do not follow our expected values. Now we will evaluate the effects of covariates on our data.

We conjecture that sales, or discounts from the original price, provide a strong impetus for people to buy and become adopters of the game. Thus, we fit the Shifted Weibull Gamma with Sales as a covariate. The following lined scatterplot and table summarize the resulting model.



N	1000000
r	3805.312
alpha	2738453
c	0.42231
B_sales	1.044664
LL Sum	-87609.9
ISM	0.546107
OSM	0.878813
TotalM	0.722765
BIC	175275.1

Figure 10: Observed versus expected (under the Shifted Weibull Gamma model with covariate Sales) number of adoptions over the callibration period

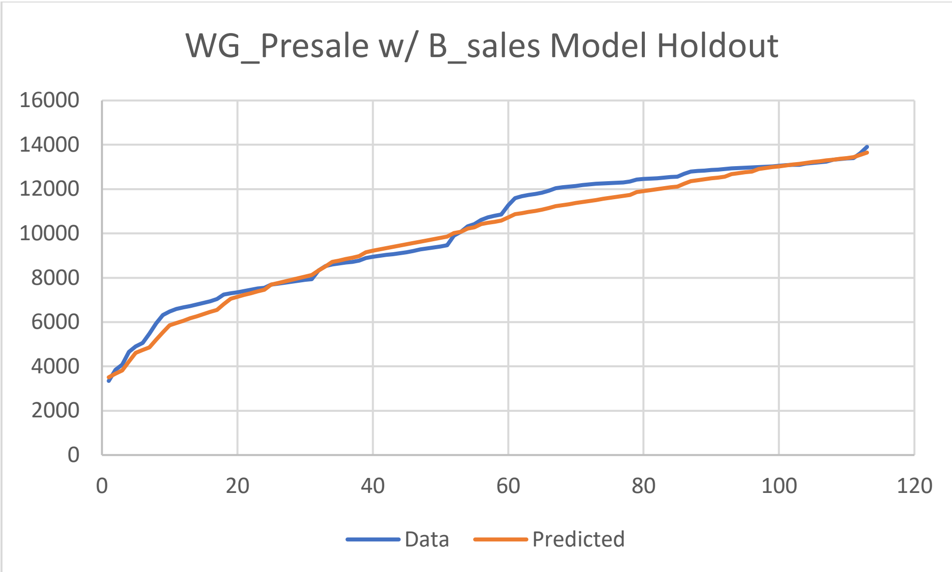


Figure 11: Observed versus expected (under the Shifted Weibull Gamma model with covariate Sales) number of adoptions over the entire period

Comparing the BICs again, we see that adding this covariate does significantly fit the data better. We also see that the gap between the out-of-sample MAPE and in-sample MAPE is decreasing, showing that the model is becoming increasingly robust. This is evidence of robustness as our model is gaining a better fit on data it is not crafted from. Another observation of significance is the degree that Sales explain the differences in the data. By adding Sales as a covariate, both the homogeneity of the population increases and the effect of duration dependence decreases. This large effect makes intuitive sense. As sales are both advertised and provide a strong incentive to purchase an item, there will be significant adoption during these periods. Furthermore, if we observe classical economics, we know that if there is more "supply" (or a lower price), more demand will be satisfied as it meets the willingness-to-pay of more consumers. While this model does fit the data better, we still see some deviation from our predicted.

To explain those other deviations, I looked at a variety of other methods of fitting models to the data, including splitting the data into segments, adding spikes, as well as utilizing more covariates. Ultimately, only the covariates for Sales of the sequel *Afterbirth* and accompanying bundles as well as releases of updates to the game made a significant improvement to the model.

It is clear why sales of the sequel *Afterbirth* affect adoption of *Rebirth* as one needs the base package *Rebirth* to play *Afterbirth.* The release and subsequent news of this additional game also generates interest into the series itself and drives adoption. As for effect of the release of patches or updates to the games, there is generally greater buzz around video games when new features are added or bugs are fixed. Reading through the extensive cult following this game claims, there is significant engagement on forums regarding this game, especially when there are hidden and often convoluted secrets buried within the updates that the community comes together to solve. In setting up my covariates, I used a simple (1/0) binary variable to indicate if there was expected any effect from the covariate. The only time I deviated from that was in setting up the covariate for Patches. As the game's creator launched a special Augmented Reality event around the game for around two weeks during the period, I doubled the effect of the patch in that time period to reflect the substantial buzz and discourse during that time period that I believe would draw in new players.

One interesting note is how the effect of the Christmas covariate was entirely negated by introducting the Patch covariate. The following lined scatterplot and table summarize the resulting models.

N	1000000
r	3972.581
alpha	2666343
c	0.399776
B_sales	0.95644
Christmas	0.428042
Afterbirth	1.179036
LL Sum	-87324.8
ISM	0.437487
OSM	0.850069
TotalM	0.656557
BIC	174732.5

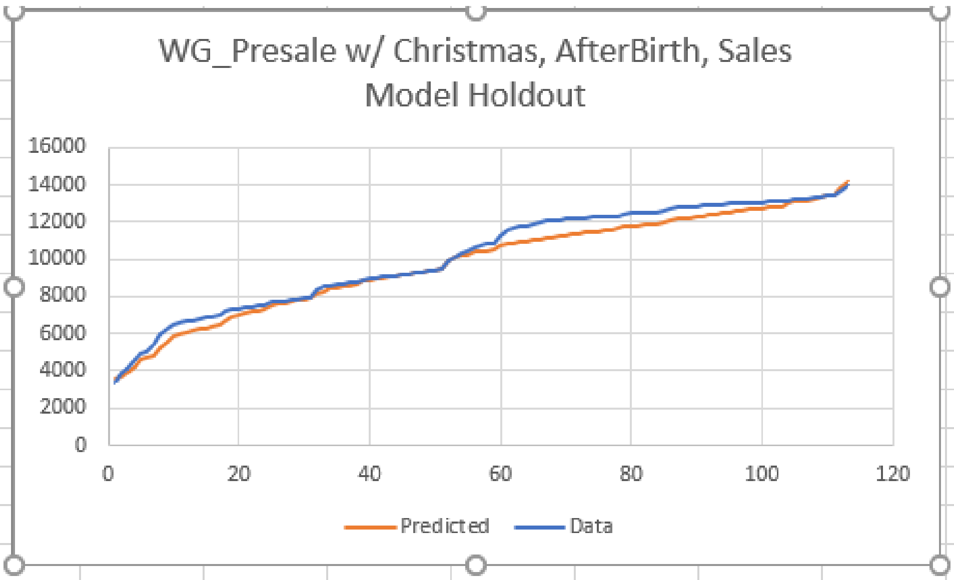


Figure 12: Observed versus expected (under the Shifted Weibull Gamma model with covariate Sales, Christmas, Afterbirth) number of adoptions over the entire period

N	1000000
r	3963.837
alpha	2743849
c	0.391644
B_sales	0.958375
Christmas	0
Afterbirth	0.752075
Patch	0.510486
LL Sum	-87211.4
ISM	0.399597
OSM	0.677305
TotalM	0.547053
BIC	174519.4

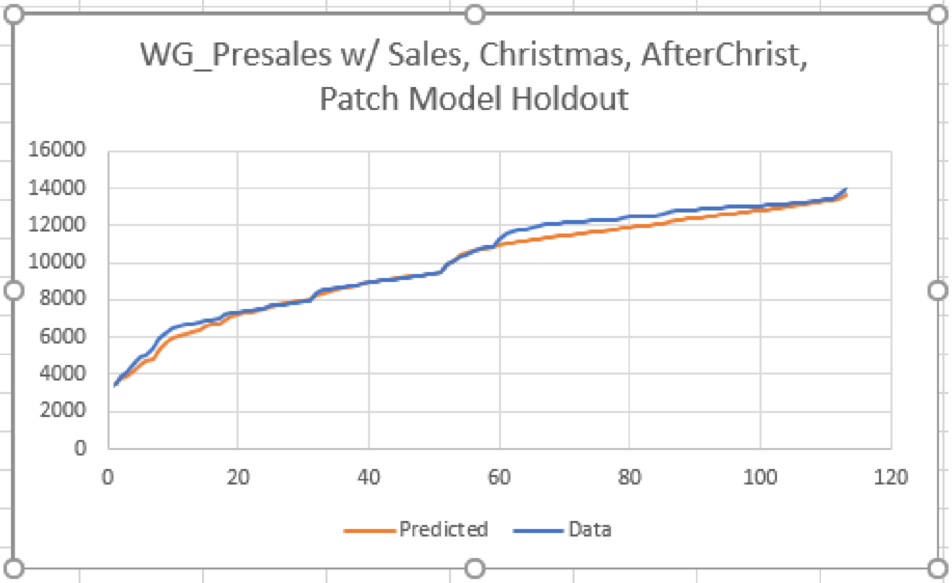


Figure 13: Observed versus expected (under the Shifted Weibull Gamma model with covariate Sales, Christmas, Afterbirth, Patch) number of adoptions over the entire period

In the model without Patch, we can see that the Christmas season contributes significantly to the fit of the model and in the model with Patch that it doesn't contribue anything. Upon closer examination, we realize that patches are usually released during the Christmas season and thus the covariates overlap and both are not necessary. I ultimately used Patch as a covariate rather than Christmas as it had a stronger effect and patches occurred beyond just the Christmas season. I also did not believe this game was of the sort that is given as a Christmas gift.

All other attempts to add additional covariates did not produce signficant positive change to the fit of the model. This leads us to our final model of the Shifted Wiebull Gamma with covariates Sales, Afterchrist, and Patches. The following lined scatterplot and table summarize the resulting models.

N	1000000
r	22860.91
alpha	15813000
c	0.391863
B_sales	0.956586
Afterbirth	0.756277
Patch	0.504384
LL Sum	-87211.4
ISM	0.400015
OSM	0.679229
TotalM	0.54827
BIC	174505.7

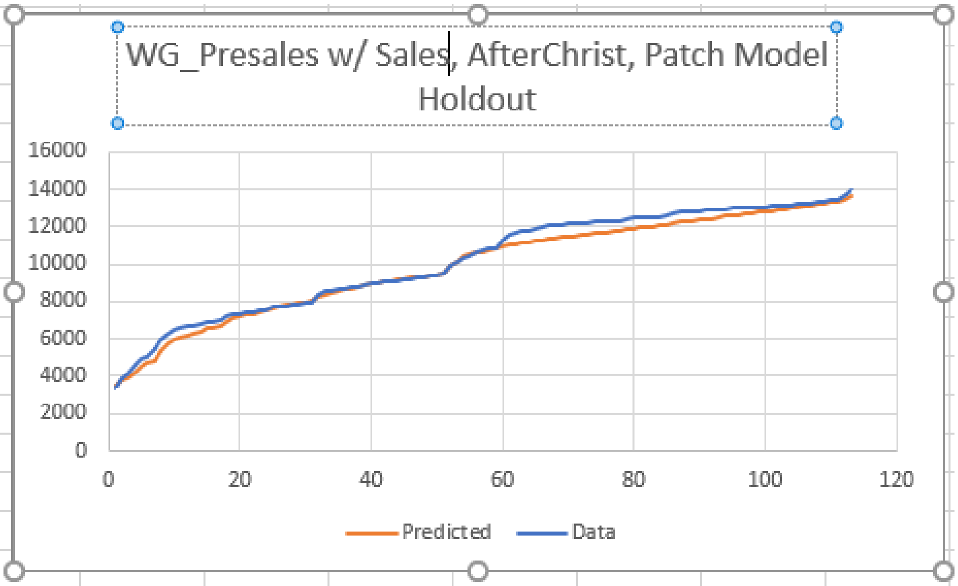


Figure 14: Observed versus expected (under the Shifted Weibull Gamma model with covariate Sales, Afterbirth, Patch) number of adoptions over the entire period

This model has the lowest BIC and there is little deviation of the observed from the predicted. The model's predicted values converge with the observed until they're nearly identical. This model's robustness is indicated by the small spread between the in-sample MAPE and the out-of-sample MAPE, with the model being only slightly worse at forecasting out of sample. With the magnitude of error quite small and decreasing, this model is well crafted to answer most managerial questions of its nature.

III Conclusions and Extension

In this paper, we have concluded with high confidence that our model is robust enough to answer nearly all managerial questions of its nature, using the data of *The Binding of Isaac: Rebirth* as an example. With the volume of demand in our dataset, our model deviates only 0.001892935 or less than 1% at its closest point and deviates only 0.024228926 or 2% at its greatest point. Given the sparse amount of data we were given to work with, this model fits quite well at an extremely computational cost.

Given more information we can expect the model to fit even more robustly. For example analyzing the social effects or impressions of news coverage on the game may yield a more effective model. As the game's founder did have strong Twitter engagement, it may also be interesting to analyze the engagement with him and how it effects the adoption of the game.

IV Index

Included are the web sources I drew on when selecting my covariates:

https://kotaku.com/the-binding-of-isaac-s-biggest-secret-nearly-broke-the-1756915840

https://isthereanydeal.com/search/?q=binding+of+isaac

https://bindingofisaacrebirth.gamepedia.com/Version_History

https://wiki.gamedetectives.net/index.php?title=Afterbirth_ARG