Box Office Revenue Forecasting: A System Dynamics Approach

Ed Ramsden
Principal, Sensorlytics LLC

Executive Overview:

System dynamics models offer tremendous potential for use as forecasting tools. The primary advantages offered by systems dynamics-based forecasting models over traditional forecasting methods include the ability to incorporate expert knowledge in the model and the ability to model highly non-linear behavior.

With advantages also come drawbacks. Because the model behavior, and resulting forecasts are dependent on expert input, it is all too easy to build a model that has little utility as a forecasting tool. To avoid this scenario, it becomes necessary for the forecaster to extensively validate a given model against historical data.

This white paper describes the development of a system dynamics model for forecasting movie box-office revenue based on early box-office returns. In addition to the description of the model we also present the results of extensive back-testing and forward (forecast) testing as an example of the degree of validation necessary for one to develop understanding of the model's performance capabilities and limitations.

1. Introduction

This white paper presents a case study outlining the development of a forecasting model based on the system dynamics paradigm [1]. While there are many techniques for building forecasting models, the system dynamics paradigm is nearly unique in that it allows the modeler to incorporate expert-level knowledge of the underlying system in the model. Most forecasting paradigms (curve-fitting, exponential smoothing, auto-regressive techniques) simply assume an underlying linear model (static or dynamic) of prescribed degree, and don't allow for the inclusion of additional knowledge of how the system behaves.

While forecasting can provide invaluable guidance in business planning and decision-making processes, too many forecasts are of dubious quality. Some of this lack-of-quality results from the inherent uncertainty of the future. Much, however, results from not examining the models used in scrupulous detail, so as to be able to justify the confidence one places in them. Some of the steps to building confidence in a forecasting model include the examination of underlying assumptions and extensive 'back-testing' of the model on historical data. Finally, once the forecaster has developed a credible model, it needs to be tested by producing forecasts of data that were not used in its formulation, and those forecasts compared to the true data. Essentially, the model needs to be held accountable for actual performance. This last item is the key ingredient missing in the forecasting technique of both tabloid psychics and a substantial number of financial-sector media pundits.

In this whitepaper, we will be using the following methodology to evaluate out models. The models will be evaluated against both the 'training' data sets used during the initial development and calibration process, as well as a 'test' data set which represents new data that has not been used in model formulation. Each of the train and test data sets contains more than 20 individual cases. While this is not large enough to provide statistically convincing results, it is large enough to provide a non-trivial degree of confidence. Finally, all of the models will be compared side-by side for performance.

2. Why the Movies?

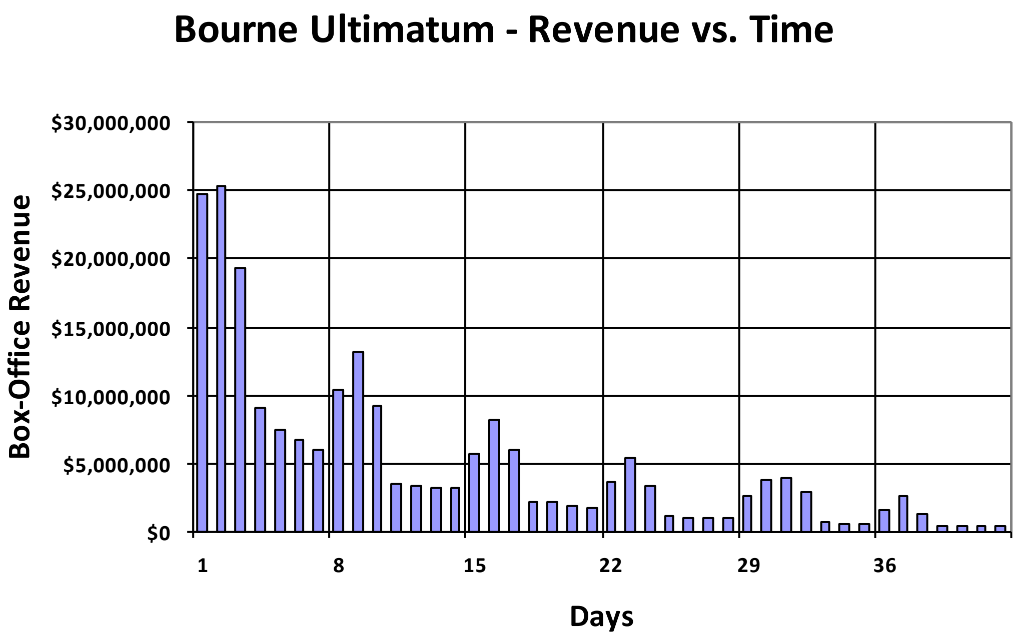
Product revenue lifecycles are a fertile area for applying forecasting techniques, if for no other reason than the potential utility of the resulting models. Although there are thousands of products for which good forecasts would be valuable, the motion picture industry has several characteristics that make it a good test-bed for model development. The first is that it produces a large number of new products with substantially similar characteristics (marketing and distribution channels, physical form, modes of consumption). The second reason is that detailed sales data are available for a large sampling of new products. For many movies, daily sales data are publicly available over a substantial portion of the movie's theatrical run [2]. This is in sharp contrast to most industrial and consumer products, for which detailed sales data are normally considered highly proprietary trade secrets.

Beyond the availability of data and its subsequent utility as a test-bed, the U.S. motion picture industry is a substantial economic sector, worthy of study in its own right. In 2007 the U.S. Motion picture industry generated approximately $9.6 billion in box-office revenues resulting from 590 theatrical releases [3]. Because of the highly speculative and risky nature of the production side of the business, it is desirable to be able to develop accurate revenue forecasts as early as possible in the product life cycle as possible. While revenue forecasts would be most beneficial at a very early stage in product development, for example when a movie concept is evaluated, forecasts at later times, such as just after release may also have value. The forecasting models developed in this white paper fall into this latter category, providing the ability to predict total lifecycle revenue based on early box-office returns (first 10 days).

3. Movie Revenue Cycle

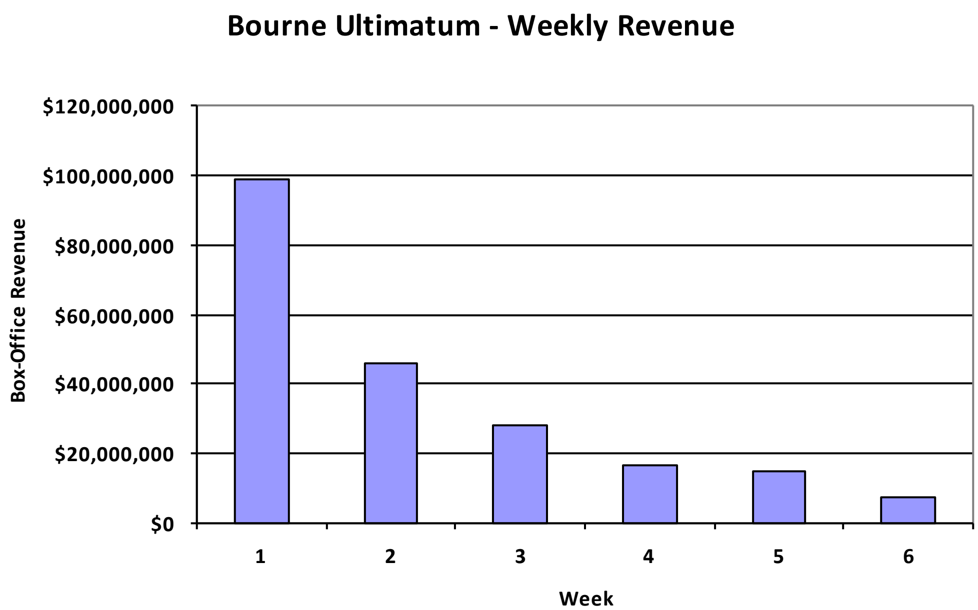
This case study considers only mass-release movies, in which a movie is typically opened nationally on more than a thousand screens. The theatrical life cycle, as measured by box-office revenue has a strong exponential characteristic with a significant variation over the course of each week, as exemplified in Figure 1 (Daily revenue for The Bourne Ultimatum).

Figure 1 - Example Daily Revenues (The Bourne Ultimatum)



While it would be possible to construct a model to deal with the day-of-the-week revenue variation, this would introduce unnecessary complications for the purposes of this case study. For this reason we will only be considering weekly revenue, where a week is defined from Friday through Thursday, as Friday is a common opening day for movies. This results in a weekly revenue chart exemplified by Figure 2. All of the movies used in this case study had Friday openings.

Figure 2 - Weekly Revenues (The Bourne Ultimatum)

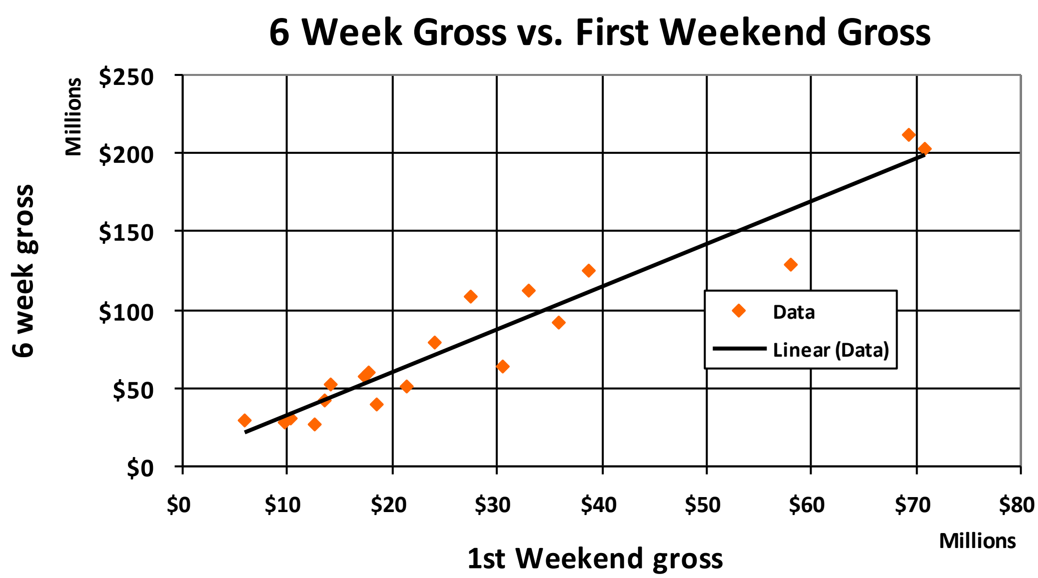


Because of complications introduced by holidays, and their associated rises in theater attendance, the motion pictures we chose for our samples were selected so that they did not have any non-weekend holidays fall within the time horizon of interest (42 days).

4. Linear Models

An obvious revenue model to build is a linear one, based on the assumption that total revenue will be proportional to early revenue. Figure 3 shows the relationship between 6 week total revenue and that of opening weekend for the motion pictures we used for our test data set. It can be seen that while the relationship is approximately linear, there is a noticeable amount of variation from the ideal (fit line).

Figure 3 - Linear Relationship between Opening Weekend Gross and 6-week Gross



As a baseline for model performance, we constructed three linear models. Each of these models forecasts total 6 week revenue as a linear function of early revenues.

Linear Model 1: 6 Week Revenue as a multiple (3.027) of opening weekend revenue:

Linear Model 2: 6 Week Revenue as a multiple (1.590) of first 10 days revenue:

Linear Model 3: Regression Model. Y = X*1.54+$1,500,000 where X is the first ten days revenue and Y is the total 6 week revenue.

The model development and testing methodology we followed in this and subsequent analyses is as follows:

1) Use a ‘Train’ data set to estimate model parameters.

2) Back-fit (retro-cast) the model to the ‘Train’ data set using the parameters obtained in step 1.

3) Compare the back-fit estimates to the actual 6 week revenue. Calculate the following statistics:
- Minimum error - the case with the most negative error
- Maximum error - the case with the most positive error
- Mean error - the average error of all cases. This is useful as a measure of model bias
- Mean Absolute Error - The average of the absolute values of the errors. This is useful as a robust indicator of model accuracy.
- Standard Deviation of error - a measure of the range of errors
- Upper and lower 2 sigma limits - minimum and maximum forecast error one might expect from the model with ~95% confidence.

4) Make forecasts on the ‘Test’ data set using the parameters derived from step 1.

5) Repeat the analysis of step 3 using the test data set results.

The above procedure was applied to the linear model cases, and the results are shown in Table 1.

Table 1 - Error Summary for Linear Revenue Models

	Proportional Model (1st weekend)		Proportional Model (1st 10 days)		Regression Model (1st 10 days)	
Forecast Error	‘Train’ Data Set	‘Test’ Data Set	‘Train’ Data Set	‘Test’ Data Set	‘Train’ Data Set	‘Test’ Data Set
Minimum	-37.6%	-31.0%	-26.7%	-22.0%	-23.9%	-23.5%
Maximum	44.7%	53.2%	26.6%	30.3%	24.8%	36.0%
Mean	5.1%	18.4%	2.0%	9.6%	1.3%	10.3%
Mean Abs. Error	18.5%	25.4%	11.7%	15.4%	11.8%	17.0%
Standard Deviation	23.9%	24.0%	14.5%	15.4%	14.2%	17.4%
lower 2 sigma limit	-42.7%	-29.5%	-27.0%	-21.1%	-27.1%	-24.4%
upper 2 sigma limit	52.9%	66.3%	31.0%	40.3%	29.8%	45.0%

One important observation of the above results is that the errors for the training data set are significantly lower than those for the test set. This implies that the models do not generalize very well to new, previously unseen, cases. Another observation is that as the amount of data used to build the model (10 days vs. first weekend only), both the training and test set errors decline.

5. System Dynamics Models

System dynamics models differ substantially from the proportional models presented in the previous section. Proportional, or linear regression models operate under the underlying assumption that the variable of interest is a linear function of some set of other ‘predictor’ variables - this does not leave much room for incorporating detailed knowledge of how the system operates beyond selecting the set of predictor variables. A system dynamics model, in comparison, absolutely requires that the modeler provide a theoretical description of the system to be modeled or forecast. While this places more demands on the forecaster, it also offers the potential for capabilities beyond those of traditional linear regression-type models.

A system dynamics model is fundamentally a differential equation-based view of reality, in which the model evolves through time as consequence of the relationships specified by the modeler. A system dynamics model is composed of three basic types of entities which must be specified by the modeler.

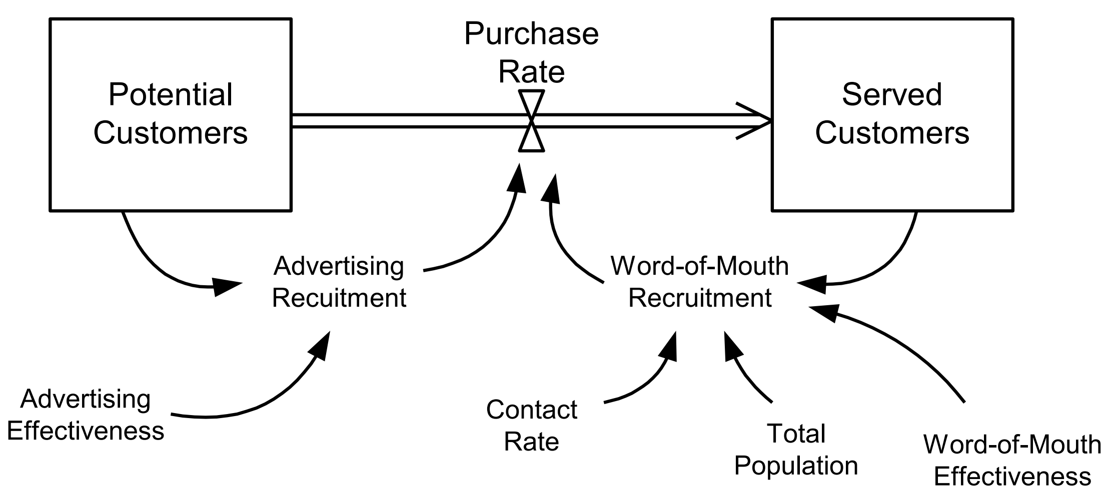
1. Stocks - these entities are variables representing quantities that vary gradually over time, such as inventory, or number of people who have seen a movie.

2. Flows - these entities define how material is transferred from one stock to another. The rate-of-change in stock levels is a function of the flows into and out of the stock.

3. Defining Relations - Flows must be defined, and can be constant, dependent on a stock level, dependent on an external variable (e.g. day-of-the-week) or dependent on functions of all of the above.

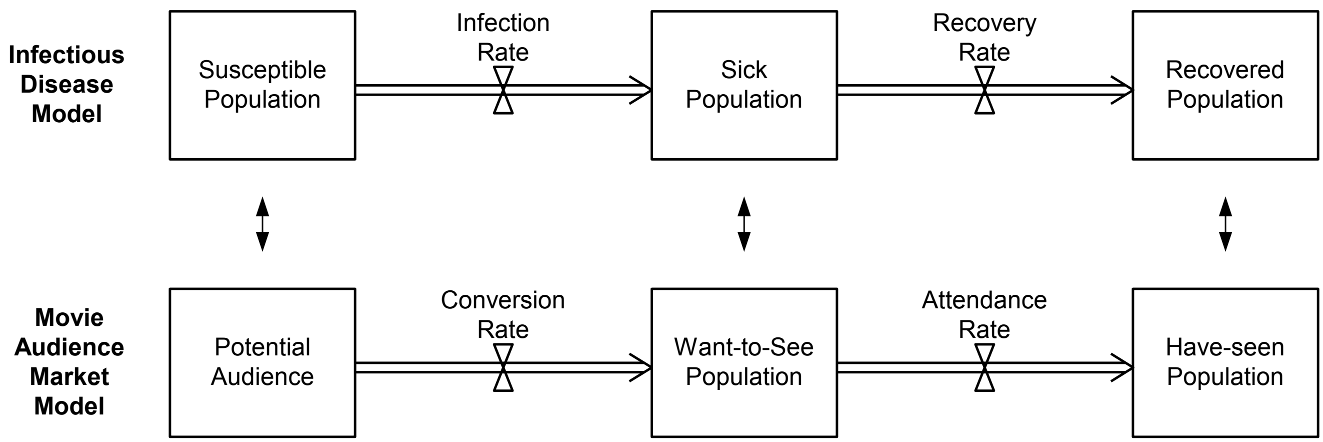
An example of a system dynamics model that has been used to model product life cycles is the Bass Model [4], shown in Figure 4.

Figure 4 - Bass Product Revenue Lifecycle Model



In this model, the two stocks are 'potential customers' and 'served customers'. A flow called 'purchase rate' transfers customers from the potential customer to served customer stocks. The rate at which this flow occurs is a function of advertising, word-of-mouth effects, and the number of potential customers. The Bass model is intended to represent sales cycles that begin slowly and build through customer word-of-mouth. While this model is applicable to the revenue cycles of limited-release movies, it doesn't accurately capture the revenue cycle behavior seen for major-release movies such as the ones considered in this case study. This is because a large base of pent-up customer demand is built up through advertising and promotion prior to the release date. For this set of circumstances, the SIR model of Figure 5, which was originally developed to model infectious disease transmission [5] is more applicable.

Figure 5 - SIR model for Epidemics (upper) and Movie-going (lower)



In the SIR epidemic model, as originally applied to disease, a 'susceptible population' is infected to become ill, based on a infection rate, and transfers to the 'sick population'. The sick population then recovers over time, transferring to a 'recovered population', who are presumed to have acquired immunity so that they can't return to the 'sick population' stock. Both infection rate and recovery rate are based on factors such as interpersonal contact, the infectiousness of the disease, the average duration of illness, and the number of people in each population.

Lane and Husemann [6] developed a variation of the SIR model to model movie attendance. The 'susceptible population' became 'potential audience', the 'sick population' morphed into an analogous population of people who wanted to see the movie, and the recovered population became the population of those who had seen the movie. One should note that although the 'meanings' of the populations and other flows are quite different in the two models, they are structurally similar.

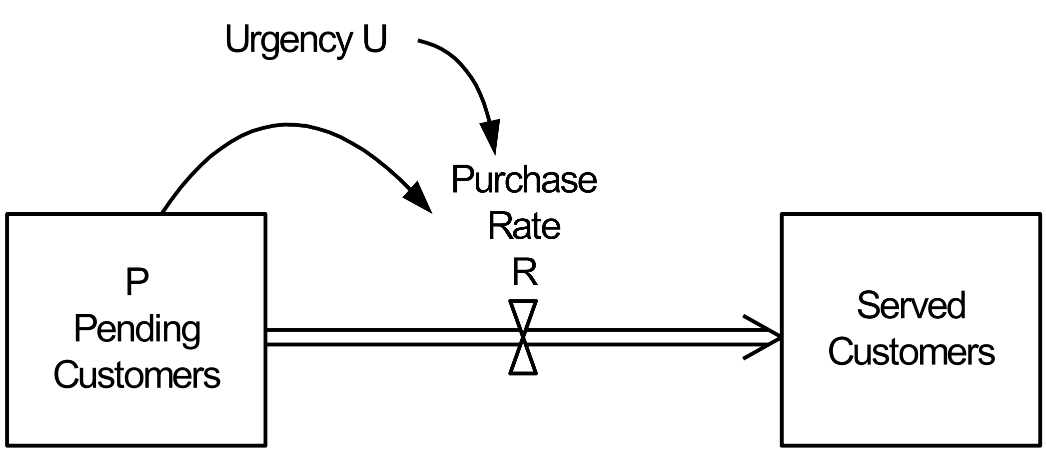
6. A Forecasting Model

In the SIR-based models of Figure 5, one may note that only the stocks and flows are shown, and that the intermediate relationships have been omitted for the sake of simplicity. Where the Bass model has 5 parameters that need to be specified, the basic SIR model requires more. While adding parameters to a model allow one to 'tweak' it to provide more accurate results when compared against known data, additional parameters can be very undesirable in a forecasting model. There are two reasons for this. The first is that the more parameters that are added to a model, the more likely one is to 'over-fit' the model to the training data. Over-fitting can result when one optimizes a model so that it does a very good job of representing the data used in its formulation. While this is not necessarily a bad thing, it can result in reducing the model's ability to accurately model new cases - a serious problem when one is formatting a forecasting model! One way to avoid over fitting is to use simple models with few 'knobs' that can be adjusted to make them fit the data.

Another reason that large number of parameters is undesirable in a forecast model is that to make a forecast requires obtaining the model parameters from very limited, early data. Beyond the obvious problems associated with how one comes up with these parameters, if one has more parameters than data from which they are derived, there exists the strong possibility of uniqueness problems - there may be numerous (or even infinite) sets of parameters that can be derived from the initial data. If there can be several possible, and equally valid, sets of parameters for a given model, how do you choose which forecast is the most valid? Minimizing the number of model parameters can minimize these problems.

For the above reasons, we are going to 'strip down' the SIR model to the bare minimum needed to replicate weekly movie-going behavior (Figure 6). We are going to ignore the effects of advertising and just assume that an initial population of pending customers exists on opening day, and that the rate at which customers buy is proportional to the number of pending customers. This corresponds to an underlying assumption that each customer has a fixed likelihood that they will go to a movie any given week - and once they have seen it will not go again. This likelihood factor is the 'urgency' factor in the model. Both the size of the pending customer stock and the urgency of seeing the movie on a given week can be estimated from early box-office data (1st 2 weekends and intervening days).

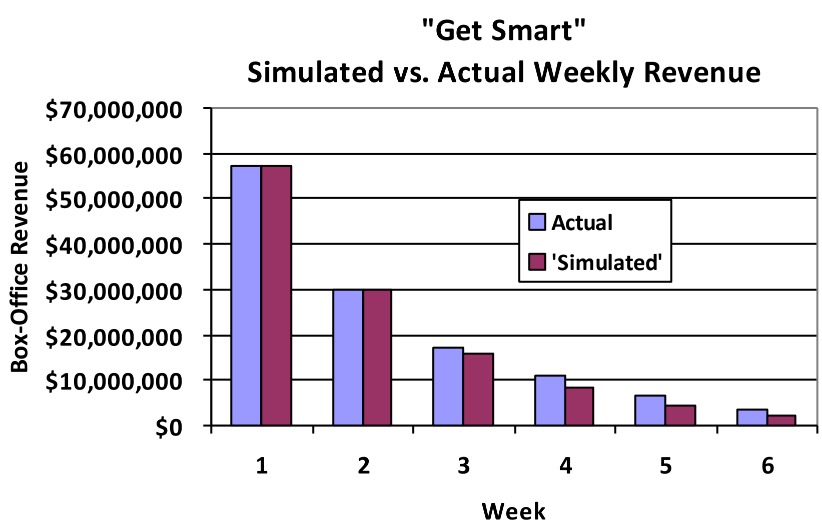
Figure 6 - Radically Simplified SIR Model



One factor that is conspicuous by its absence from this model is word-of-mouth feedback. While certainly a major consideration in how well a movie does commercially, this is something that we presently can't readily infer solely from early box-office return data, so it is best not considered for a first model.

When one runs a systems dynamics model, whether on a spreadsheet or using specialized system dynamics simulation software such as Vensim [7], one gets results such as those of Figure 7, which are shown along with the actual box-office returns.

Figure 7 - Simulated vs. Actual Box Office returns (Get Smart)



If we repeat the procedure used to evaluate the linear models previously discussed, this results in the data shown in Table 2. Note that the mean error, mean absolute error, and standard deviation of the test data set results for the system dynamics model is noticeable lower than those of the linear models listed previously in Table 1.

Table 2 - Forecast Errors from System Dynamics Model

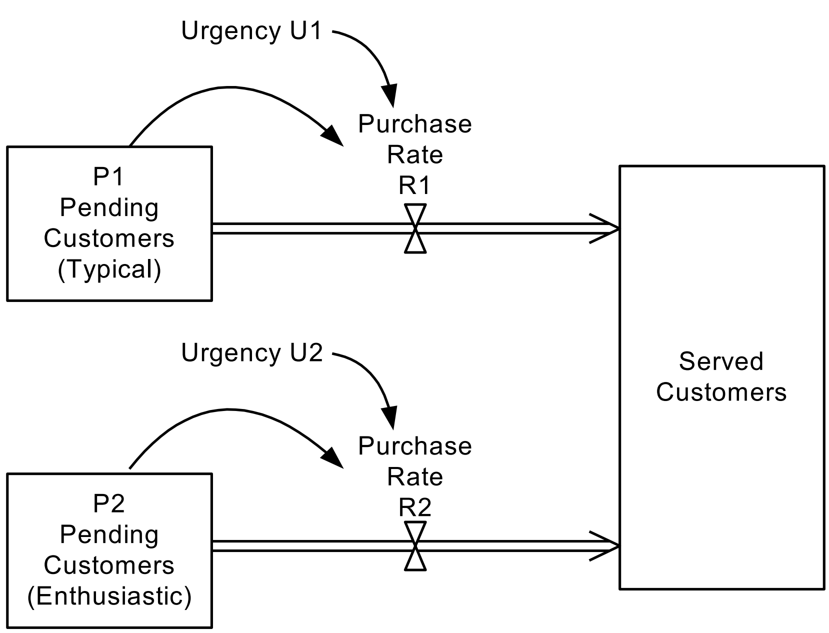
Forecast Error	'Train' Data Set	'Test' Data Set
Minimum	-12.5%	-14.5%
Maximum	18.7%	12.3%
Mean	-2.0%	-4.5%
Mean Abs. Error	7.9%	7.5%
Standard Deviation	9.3%	7.7%
lower 2 sigma limit	-20.5%	-19.9%
upper 2 sigma limit	16.6%	11.0%

7. Can We Do Better?

It may be possible to increase the forecast accuracy of the model by adding complexity. The down side of adding complexity is that adding more entities requires the estimation of more model parameters. Another issue in adding complexity is hypothesizing credible effects to add.

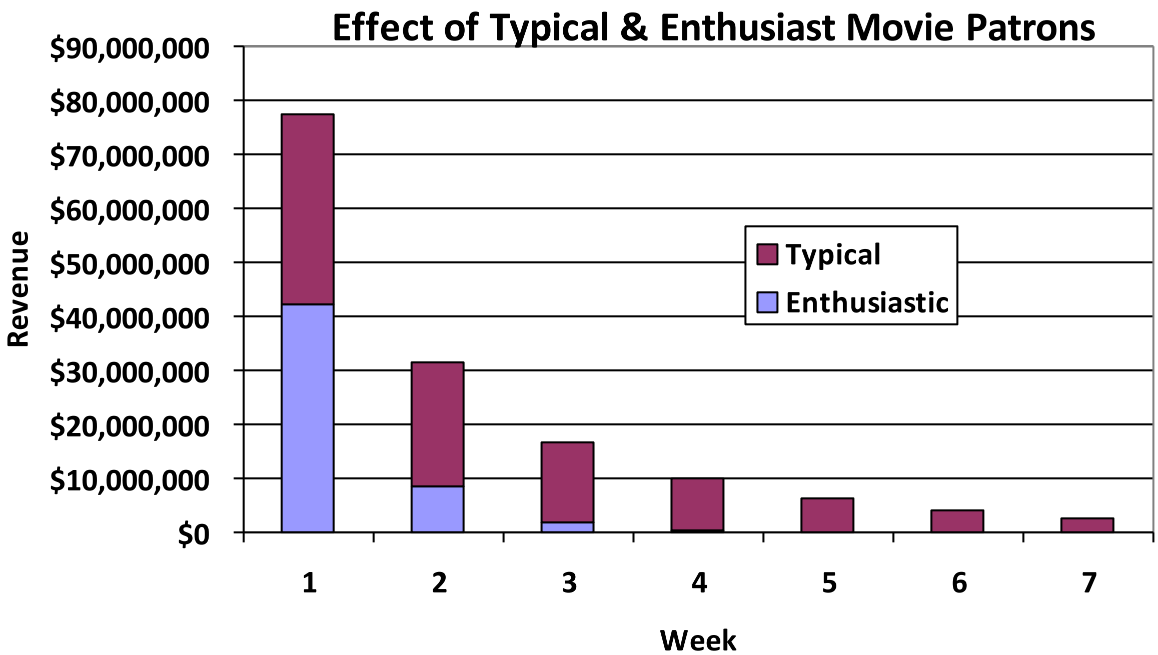
While word-of-mouth is undeniably an important effect in movie attendance, we hypothesize an alternate effect that may (or may not) come into play. In the previous model, we assumed that moviegoers are a single homogeneous population, all of who have the same level of urgency to see a given movie. Observation suggests that this is not always true. When one hears reports of people camping overnight on sidewalks to be the first to get to see a 'blockbuster', one suspects that different individuals experience widely ranging levels urgency when it comes to seeing a movie. This suggests that a model with two separate customer stocks, such as shown in Figure 8.

Figure 8 - Two-stock Model of Customer Demand



In this model we assume distinct stocks of 'Typical' customers and 'Enthusiastic' customers. If the urgency rates for the different stocks differ, one would expect to see the behavior shown in Figure 9, where the urgent customers dominate the first week's attendance, but the typical customers come to dominate later weeks.

Figure 9 - Effects of Combination of Typical and Enthusiastic Movie-goers



The price of this added complexity is the need to estimate four model parameters instead of only two (initial conditions for each customer stock and urgency rates). Because of data limitations, and from observations of the results of the first system dynamics model, we will assume the two urgencies (U1 = 0.35, U2 = 0.65). By assuming these two values, we only have to estimate the two initial populations. For this model, we followed a somewhat different procedure for the 'Train' data set, where we fit the model to all 6 weeks of known data, as opposed to just fitting the model to the total final revenue. This resulted in a very good back-fit compared to any of the previous models. We then performed the 'Test' data forecasts in the same manner as we did with the previous models, except that the model parameters were estimated from the first 14 days of data instead of the first 10 days. This was done to simplify the parameter estimation process (using two full weeks vs. just two weekends). Note that the mean error of the test data set is noticeably better than that of the original system dynamics model, but that the mean absolute error and standard deviation are nearly identical. This corresponds to a lower degree of bias in the forecasts made from the two-stock model, but a similar degree of accuracy (uncertainty).

Table 3 - Two-stock System Dynamics Model Errors

Forecast Error	'Train' Data Set (6 week fit)	'Test' Data Set (2 week fit)
Minimum	-3.0%	-10.5%
Maximum	3.0%	18.3%
Mean	0.0%	-1.4%
Mean Abs. Error	1.4%	6.3%
Standard Deviation	1.8%	7.6%
lower 2 sigma limit	-3.6%	-16.6%
upper 2 sigma limit	3.6%	13.9%

8. Comparison of Models

In this section we compare the performance of all of the models presented to this point. Figure 10 shows the mean error vs. forecasting method (model type), while Figure 11 shows the standard deviation vs. forecasting method (model type). As previously noted, this metric indicates the systematic bias of a given model. One can see that the two system dynamics models have a noticeably lower degree of bias (both for train and test data) than the linear models.

Figure 10 - Mean Error vs. Forecasting Model

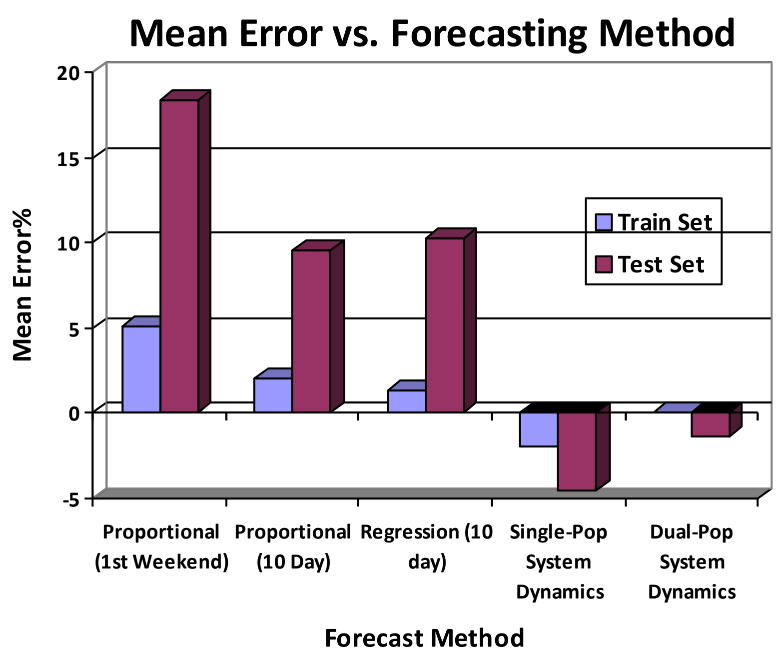
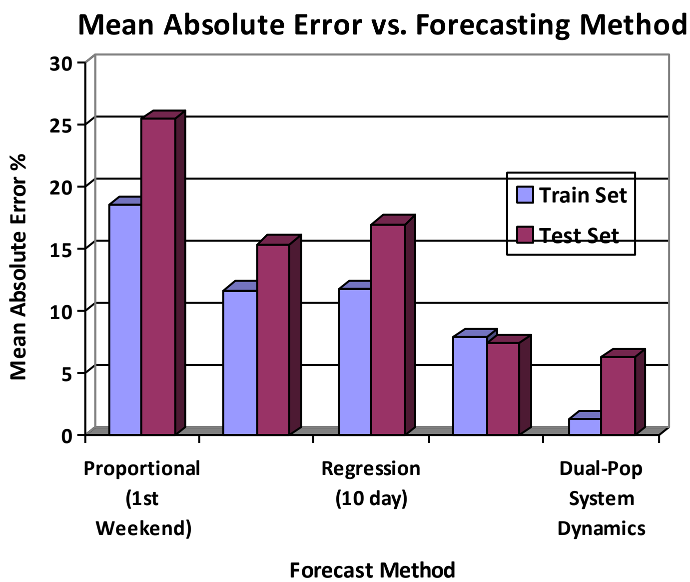


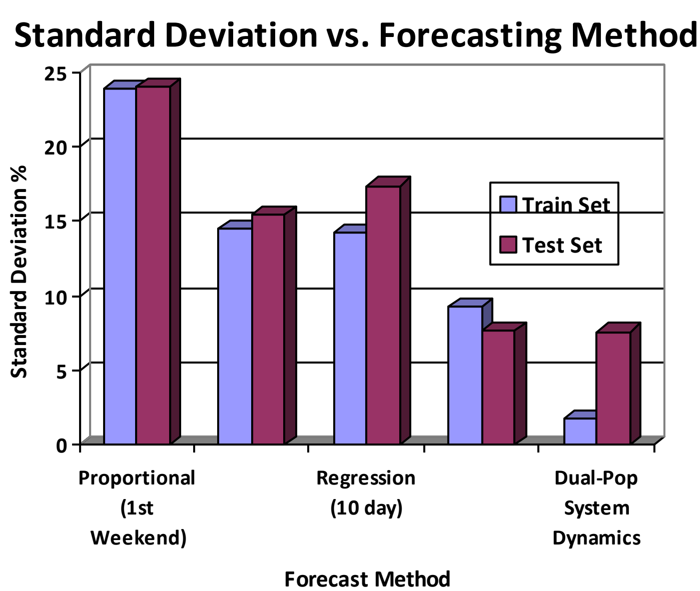
Figure 11 shows the mean absolute error (MAE) for each model with both train and test data. This metric reflects the accuracy of each model's predictions. The system dynamics models both have much lower errors, and higher accuracy, than the linear models. One important point to note, however, is that while the MAE for the two-stock model's performance on the training data set is much lower than that of the single-stock model, its performance on the test data set is comparable. This is an indicator that the additional complexity is not contributing much to the ability to provide accurate forecasts.

Figure 11 - Mean Absolute Error vs. Forecasting Model



Finally, Figure 12 shows the standard deviation of the errors in model forecasts. The magnitude of the standard deviations are comparable to the MEAs for Figure 11. This suggests that the models are not being strongly affected by the presence of outliers, which tend to increase the standard deviation more than MEA.

Figure 11 - Standard Deviation of Error vs. Forecasting Model



9. Conclusions

This white paper has presented a case study of using system dynamics models for market revenue forecasting. While focusing on the motion picture industry, models of a similar nature may be developed for other industries where a causal market structure can be established. The emphasis of this case study was the extensive use of data for validating model performance, and the comparison of the highlighted models to other models to be used as performance benchmarks. One key point of this study is that more model complexity is not necessarily valuable. While the use of more complex models may result in the ability to better fit the model to previously observed data, it may not result in higher quality forecasts. This was shown through the development and validation of both a simple and a more complex system-dynamics model.

Acknowledgements

I would like to thank Dr. Wayne Wakeland of the Department of Systems Science at Portland State University for his input and contributions to this analysis through numerous reviews and discussions of the models and results.

References

[1] Roberts, N., Anderson, D., Deal, R., Garet, M., Shaffer, W., *Introduction to Computer Simulation: A System Dynamics Modeling Approach*. Addison-Wesly, Reading, Massachusetts, 1983.

[2] www.the-numbers.com - web site with box-office and other information on a wide range of movies.

[3] Entertainment Industry Market Statistics 2007, MPAA, retrieved from www.mpaa.org/USEntertainmentIndustryMarketStats.pdf, May 8, 2009

[4] Sterman, J., *Business Dynamics: Systems Thinking & Modeling for a Complex World*, pp.332-339, Irwin McGraw-Hill, Boston, 2000.

[5] Ibid, pp.303-309

[6] Lane, D.C., Husemann, E., *Movie Marketing Strategy formulation with System Dynamics: Toward a multi-disciplinary adoption/diffusion theory of cinema-going*, retrieved from http://www.systemdynamics.org/conferences/2004/SDS_2004/PAPERS/250LANE.pdf, May 8, 2009.

[7] Vensim is a simulation tool developed by Ventana Systems, www.vensim.com

About Sensorlytics

Marketing, Communication and Analytic Services for the Sensor and Technology Industries

Sensorlytics LLC is a consultancy that provides services to help sensor and technology companies achieve their marketing goals. Our primary service offerings include:

Market Research & Analysis
Technical Communications
Product Definition
Analytic Services

For more information on what we do, the services and resources we provide, and how we might be able to assist you, please visit our website at:

www.sensorlytics.com