Design influence on player retention: A method based on time varying survival analysis

Thibault Allart, Guillaume Levieux, Michel Pierfitte, Agathe Guilloux, Stéphane Natkin

Design Influence on Player Retention : A Method Based on Time Varying Survival Analysis

Thibault Allart[*†§], Guillaume Levieux[†], Michel Pierfitte[*], Agathe Guilloux[§] and Stephane Natkin[†]

[*]Ubisoft Production Internationale
126 rue de Lagny, 93100 Montreuil-Sous-Bois
[†]Conservatoire National des Arts et Metiers
292 rue Saint Martin, 75141 Paris
[§]Universite Pierre et Marie Curie (Paris 6)
4, place Jussieu, 75252 Paris

Emails: thibault.allart@gmail.com, guillaume.levieux@cnam.fr , michel.pierfitte@ubisoft.com,
agathe.guilloux@upmc.fr , stephane.natkin@cnam.fr

Abstract—This paper proposes a method to help understanding the influence of a game design on player retention. Using Far Cry[®] 4 data, we illustrate how playtime measures can be used to identify time periods where players are more likely to stop playing. First, we show that a benchmark can easily be performed for every game available on Steam using publicly available data. Then, we introduce how survival analysis can help to model the influence of game variables on player retention. Game environment and player characteristics change over time and tracking systems already store those changes. But existing model which deals with time varying covariate cannot scale on huge datasets produced by video game monitoring. That is why we propose a model that can both deal with time varying covariates and is well suited for big datasets. As a given game variable can have a changing effect over time, we also include time-varying coefficients in our model. We used this survival analysis model to quantify the effect of Far Cry 4 weapons usage on player retention.

I. Introduction

The amount of data collected by game companies about how their games are played is constantly growing. Indeed, consoles and PC are almost always connected to the internet, allowing Game Studio to virtually track every player interaction with the game. However, interpreting this massive amount of data is still a key challenge for game companies.

One of the goal pursued by game developer is not only to have many players, but to have each of them playing a lot. Players should spend more time than just trying the game for a few minutes : they should really enjoy the game and explore the whole content developers spent a long time to create. Having many players leaving the game in the first hours could really be considered as a design failure.

Game studios are thus often analyzing a game's *player retention*, that is to say, the proportion of remaining players after n hours of playtime. In this paper, we focus on using statistical analysis of tracking data to improve player retention. We need to help the designers to identify the design elements that seem to maintain players inside the game, and those who don't. By nature, game design is an incremental process and

games need to be tested to evaluate if players have fun playing it [17].

A lot can be learned by only monitoring playtime, including detecting when a player is about to leave the game. For instance, an issue with the tutorial could be highlighted by a player retention drop in the first minutes of the game. However most thing are not as easy to identify. An unbalanced weapon will have an unnoticeable impact in the quit rate curve, even if it causes lots of players leaving the game. Different players will use the given weapon at different times in the game leading to a wide spray on the time scale. A good way to detect such impact is to use statistical modeling. In order to inform designers on which modifications can be done, we need to investigate how in-game metrics are related to retention.

One of the most interesting aspects of video game tracking systems is that they allow us to get a large amount of accurate data on player behavior. In contrast, in the field of health care, the number of individuals is closer to hundreds than millions. Furthermore, in longitudinal studies, the patients might be monitored every week. In video games, we can track any action anytime in the virtual world. This deep tracking can help us to better understand the design's link with player retention but we have to adapt our model to be able to handle such a high amount of data.

II. Contribution and main result

This paper proposes a method to extract actionable insight from tracking data. There exists many black box models, e.g. random forests or deep learning, that may be used to predict player departure, but, as our goal is to give actionable insight to designers, we are restricted to interpretable models.

First, we describe how a simple metric like playtime can be used to identify time periods when players stop playing the game. We also show that this simple metric can be calculated on publicly available data, and thus can be used by anyone to realize a benchmark and see how a given game differs from concurrent ones. Indeed, online gaming platforms, like Steam, are gathering data on their players, and some of these data

can be retrieved by anyone. This means that any researcher can nowadays analyze high level data on gamer's playing habits. However, this data are often limited to the time spent by players in each game, and do not provide any insight on the player's detailed behavior within the game.

Then, we develop how to model the link between the player's behavior and the player's departure. Thereby, we introduce a well-known survival analysis model and show how it can be used with constant-over-time data. Such model works well for covariates that don't change over time. However, using this model on time-varying data, which is often the case in video games, can lead to wrong conclusions.

Finally, we detail the complexity of time-varying data. We show how information can be visualized and explain why we need to take time-varying covariates into account. We introduce a new algorithm that is able to both analyze a high amount of tracked data and to take into account time-varying covariates. We show how it can be applied to analyze Far Cry 4 weapon usage, and what kind of design recommendations can be done.

III. RELATED WORKS

A video game is an often-complex interactive multimedia system. As such, it can be described by many quantifiable variables, depending on the *gameplay* it provides as well as, for instance, technical aspects. Gameplay is a relationship between the player and the game system and thus gameplay variables may describe part of the game system, e.g. a gun's rate of fire, as well as the player's behavior, e.g. the preferred weapon. Technical variables can describe the position of a HUD's icon to the average amount of network lag. Many of these variables can be modified by the development team, and are often called game features when related to the gameplay. We will use the more abstract term of *game variables*.

Game analytics is a wide area including visualization [18], [4], clustering [2], [21] and prediction [12], [20]. In this work we focus on ways to understand the link between game variables and the risk that players quit the game. This topic has already been treated by some authors.

Weber[19] used a regression model to determine which game variable might have the most influence on player retention. In this research, Weber et al. study the case of an american football game in which the player has to perform multiple matches. Instead of directly predicting the game duration, the authors use the number of matches played as a target variable. Our approach extend their work for cases where we do not study a number of matches but a more general duration measure, as playtime.

Harrison et al[10] used N-gram models to dynamically attribute quests to players. They are thus able to select a sequence of quests that players are more likely to accept in order to increase player retention. However such models present an exponential complexity with the number of different actions available for the player, and might such be inapplicable to datasets with many game variables.

The largest study on playtime distribution was done by Bauckhage [3] on approximately 250 000 Steam players. They concluded that playtime seems to follow a Weibull distribution. However, using Ubisoft® tracking data, we observe that Ubisoft games seem to follow a log-Normal distribution. This can be explained by differences in the tracking system. Furthermore, playtime distribution of free-to-play games is quite different from paying games due to a high proportion of *droppers* (players who played only one day). Knowing that, using a parametric regression model (based on density distribution) could not be adapted to every games. We solve this problem using a semi-parametric approach.

Survival analysis has already been used in video game analysis. Chen[5] use survival analysis to quantify the effect of network quality on player retention in Online Games. They found that both network delay and network loss have a link with player retention. We propose to extend such work by including time-varying information in the model.

IV. METHOD

In this section we propose a method to analyze a video game's retention, starting from a general, simple approach to a more complex modeling of the link between game variables and player retention.

First, we propose to analyze playtime and compute a *hourly survival curve*, that can both be used to get a first insight on the game's retention as well as to compare it to other games with publicly available data. Then we introduce a way to quantify the link between a given game variable and player retention, namely the Cox model. After what we point out that a lot of game variables change over time. This temporal evolution has to be included in the model, otherwise it can lead to wrong results. We propose a model that includes time-varying covariates in Cox model and can deal with huge dataset.

A. Playtime Analysis

Playtime is one of the most important metric in video game analysis. First, playtime allows us to know when the players stops playing. From this, we can derive and represent metrics like *retention*, *quit rate* and *hourly survival*.

1) Retention: Retention measures how many players remain in the game after a given time. We focus here on playtime retention, but it can be computed using different time measure (lifetime, played days, ...). Retention is represented as players survival curve. Let T being the random variable of time at which a players leaves the game, then we have :
$S(t) = \mathbb{P}(T > t)$

If we want to know when we lost the most players we would better look at quit rate.

2) Quit Rate: The hazard function, also called quit rate in video game industry, gives us the instantaneous rate at which players quit the game :

$$h(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \qquad (1)$$

Retention and Quit Rate only needed a playtime measure to be computed.

A locally high value of quit rate may indicate a problem in your game.

3) Hourly survival: Using a one hour binning, quit rate can be interpreted as how many players stay from one hour to the next one. We call this *hourly survival* in the next section. This metric is less accurate than quit rate and is useful when we don't have a very precise measure of playtime, as it is the case in our benchmark data.

B. Benchmark

Retention metrics can be analyzed both in an absolute and relative fashion. First, in an absolute fashion, any inflection in the curve might reveal a design problem and be investigated. But from a relative point of view, comparing the game's retention curve to other similar game curves can help to understand where the game strengths or weaknesses are. It is hard to tell whether a 85% first hour hourly survival is good, but if at some point, the game is much worse than many other similar games, then there might be a specific, hopefully fixable issue.

However, to perform such a comparative analysis, we need to know about many games playtime information, and many game companies do not share this kind of information. Hopefully, Valve's Steam online gaming platform provides an A.P.I that allow developers to query part of Steam users information, such as playtime information about the game titles owned by a specific player. Such an API may be used to regularly gather playtime information from a random set of players and then compute playtime statistics, as did for instance Bauckhage et al [3].

Moreover, this approach is used by the website SteamSpy [9] to give an approximation of playtime distribution for every Steam games. Thus, if we want to get high level information about playtime for Steam games, we do not need to directly query the Steam API but can rely on SteamSpy data instead. Currently, SteamSpy do not provide any API but data can easily be retrieved manually from the website.

Of course, data from SteamSpy are less accurate than internal data from the in game's tracking system. That is why we prefer to perform an hourly binning and thus look at the hourly survival curve in that case. Each value is the percentage of players who played for at least n hours and will continue playing for at least $n+1$ hours. For instance we have computed the hourly survival curve for some of the most played games on Steam, based on SteamSpy data (fig. 1).

These curves can be interpreted as follows : among Counter Strike Global Offensive players who have played at least 4 hours, 98% of them have played at least 5 hours. Which means that 2% of the players who played at least 4 hours have left the game between 4 and 5 hours of playtime.

We can see that Dota 2 and Team Fortress 2 have around 80% of hourly survival in the first hour. That means that around 20% of players leave the game in the first hour of playtime. We may hypothesize that as both these games are free to play games, much more player will test them than if they had to pay for it. We may also speculate that these player
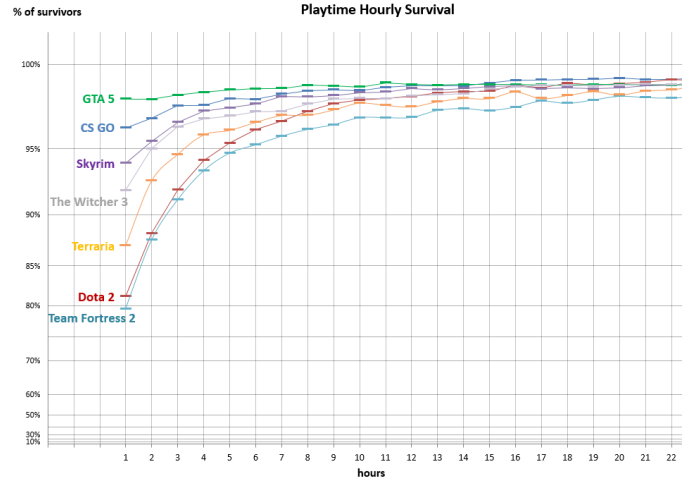

Fig. 1. Hourly survival of some games available on Steam using a logarithmic scale

have more chance to stop playing the game, thus leading to a low hourly survival during the first hours of the game.

On the contrary, Grand Theft Auto V has an exceptionally flat hourly survival curve. The game succeeds in retaining more than 98% of players hour by hour.

Looking at hourly survival curve and Quit rate may help to identify temporal windows where lots of players quit the game. In linear games it can be linked to a given game level. However, in non-linear game such as open-world games where the gameplay leaves much more freedom to the player, it is often impossible to relate a temporal measure like playtime to a spacial or a progression measure like player's position or player's progression in the missions.

Playtime analysis is thus inherently limited, and while it is a valuable first step, one will want to include other variables in the analysis and build a more thorough model.

C. Cox Regression

There are many ways to model the effect of covariates on a given metric. In our case we are interested in the retention, which is a duration value. Modeling the player retention can be done by performing a *survival analysis*. One of the most used model in survival analysis is the Cox proportional hazard [6]. It considers that every player has a given risk to quit the game at time t. This risk depends on each player's individual characteristics, represented by a set of p game variables $(X_1, \ldots, X_p)$ in the model. A player characteristics can be considered as a quantifiable variable, that can be measured for any player, and that gives us information about a game variable we want to study. For instance, if we want to know if the AI is not too strong on close combat, which might be the game variable "AICloseCombatDamage", we will calculate the player characteristic "NbDeathInCloseCombat", counting the number of deaths for each player while in close fight with an AI. The β_j related to "NbDeathInCloseCombat" might then help us tune the AI with regard to retention. We note $X_{i,j}$ the value of the covariate j for a given player i. Coefficient β_j
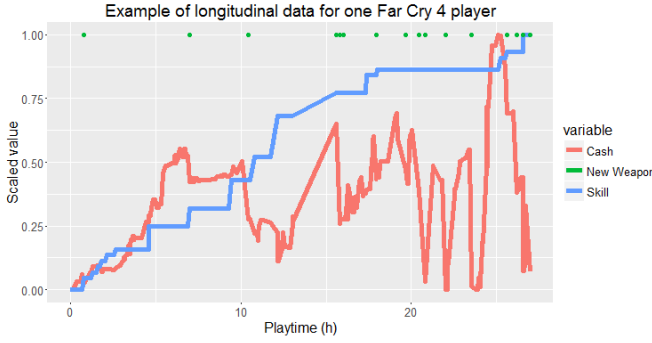
Fig. 2. Example of longitudinal data for one Far Cry 4 player. Blue curve represent the number of skills unlocked by the player. Red curve is the amount of cash own by the player. Green point indicate that player get a new weapon.

quantifies the effect of the covariate j on the hazard rate. The risk that this player stops playing at a given time t is modeled as

$$\lambda(t|X_i) = \lambda_0(t) \exp\big(\sum_{j=1}^{p} X_{i,j}\beta_j\big) \qquad (2)$$

This model is called proportional hazard because the quit rate function λ of each player is proportional to a baseline quit rate function $\lambda_0(t)$.

Cox survival analysis will help us create a model of game variables on player retention. However, we must take into account the fact that while the player gender or age are constant, many game variables that we want to study vary a lot during the game. In a shooter game, the player does not use the same gun during the whole game. In a Role Playing Game, character skills change over time and the way a player uses them vary over time. Our model needs to take into account such time-dependent covariates.

D. Time Varying data

Covariates may change very often during a game session, like the number of enemies the player has killed, the amount of money he has won or spent, or his position in the world. Each player characteristic can thus be considered be as a time series curve (fig. 2).

A classical way to deal with time-varying data would be to add new covariates that try to summarize them, such as the $mean(x(t))$, $var(x(t))$ or to decompose them as basis functions (Wavelets).

However, summarizing time-varying covariate leads to a loss of information, and basis decomposition leads to a loss of interpretation. Furthermore, applying a model made only for constant data on aggregated time varying data leads to a scaling problem. Metrics of a player who played 5 minutes will be compared to those of a player who played 80 hours. One can try to create a ratio by dividing game variable by playtime, but often metric evolution over time is not linear in time. In fact metric evolution is related to design and game variable cannot simply be compared at different time.

To avoid all these problems we introduce a model that deals with time varying game variables $X(t)$. For almost the same complexity cost we also allowed the covariates effect to change over time $\beta(t)$, meaning that the same element can have various effect depending on time.

E. Modeling Time varying data

Extensions of Cox proportional hazard have been proposed to deal with time-varying covariates and coefficients [14].

$$\lambda(t|X_i) = \lambda_0(t) \exp\big(\sum_{j=1}^{p} X_{i,j}(t)\beta_j(t)\big) \qquad (3)$$

However current implementations of such model are based on matrix inverse and iterative kernel smoothing that makes the model unable to deal with huge datasets. We propose a piece wise constant model that makes the minimization problem separable in the number of individuals. This allow us to use a stochastic gradient descent algorithm as Adagrad [7] or Adadelta [22] to solve the minimization problem. Technical part and theoretical guaranties are developed in [1]. Stochastic gradient algorithm allows us to load only few datapoints in main memory and thus to analyze datasets that are bigger than RAM capacity.

Our model has been implemented in a C++ library and interfaced with the R programming language [15] thanks to RCpp [8], to facilitate it usage by game analysts.

As it is the case with Cox proportional hazard, coefficient can be directly interpreted as an effect on retention. Time-varying coefficients cannot be summarized as a table of scalar value. Instead we need to plot the $\beta(t)$ curves. A positive value of the coefficients means that the game variable related to this coefficient has a positive link with quit rate, meaning that player has higher chance to quit the game in a small time interval. Each curve can be interpreted as follows : for a given time t, if we set all features value to a constant except the one of interest $X_j(t)$, a one point increase in the feature value $X_j(t)$ leads to multiplying the chance that players will leave the game in a short period of time by $\exp(\beta_j(t))$.

V. DATASETS

As Far Cry 4 is a shooter game, weapon usage can be considered as part of its core gameplay. Thus, we will focus our analysis on weapon usage but the method can be used with any covariates that may be tracked, time-varying or not.

A. Far Cry 4

Far Cry 4 is a first person shooter in which the player explores an open world named Kyrat. The game was released on November 18, 2014. During the game, the player discovers new weapons. He can carry four of them and switch between them to modify his strategy. We consider that weapon selection is highly related to play style and can thus give a lot of information on which part of the gameplay is experienced by the player. A player who wants to play stealthily could approach a camp using a long range weapon, as a sniper rifle equipped with a silencer, then a Bow, and finally get to his

objective using a short range silenced weapon. Another player could choose to fly over the camp with an helicopter armed with a one hand rocket launcher, shoot three or four rockets over the main guarded zone before landing inside and draw his shotgun to complete the mission, which will be a completely different play style and thus related to different part of the gameplay Far Cry 4 has to offer. As a result, if rocket users have a very different retention than silencer users, then it might give us a useful insight on which part of the gameplay the developers should modify.

B. Metrics description

Far Cry 4 tracking system monitors the number of kills made with the following weapons :

- Assault including AK-47, STG-90, F1, MS16, P416 and A52
- Auto Crossbow
- Bait. Player can launch meat to bait animals who can attack enemies. The kill is then attribute to player.
- Bow including Recurve Bow and Hunter Bow
- Grenade
- Knife thrown
- Light Machine Guns including PKM, U100, MKG, MG42 and DshK
- Machete. Available only in Shanath Arena.
- Molotov cocktail
- Pistol including Mark IV, M-712, 1911, 6P9, A.J.M. 9, D50 and .44 Magnum
- Rocket launcher including M79, RPG-7, GL-94, GL-A87 and LK-1018
- Shotgun including D2, M133, 1887 and SPAS-12
- Submachine Guns including A2000, MP5, A99, Slorpio, Vector .45 ACP, BZ19
- Sniper rifle including SVD, M-700, SA-50, Z93

Others game variables related to killing are :

- Alert : The number of kills made with AI in alert detection state.
- Animals : The number of animals killed by the player.
- Cautious : The number of kills made with AI in cautious detection state.
- Combat : The number of kills made with AI in combat detection state.
- Distance Kills : The number of kills made over 40 meters.
- Close Kills : The number of kills made under 40 meters.
- Headshot : The number of Headshot kills.
- Idle : The number of kills made with AI in idle (undetected) state.
- Silencer : The number of kills with a silencer equipped.
- TagCamera : The number of kills with enemy previously tagged with camera

C. Visualizing time varying information

As metrics change over time for a given player, it is not appropriate to summarize a given covariate in terms of mean or standard deviation. In this case it is more informative to show how a given metric changes over time. We can represent
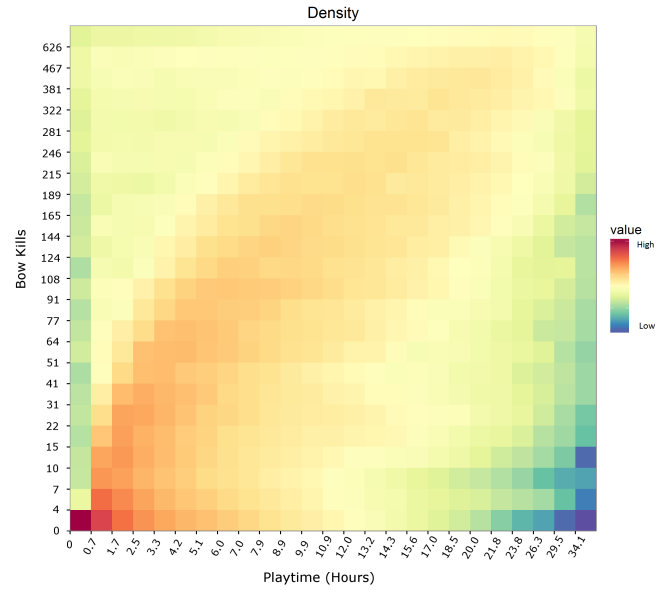

Fig. 3. Density of number of kill using bow over time

the time variation of a metric as a heatmap of players density (fig.3).

As different users can have a different tracking frequency in database, we have to be cautious. Each square of the heat map represents the number of player having the corresponding covariate value at the corresponding time. Thus we need to ensure that we take into account only one observation by player, if he is still playing at that time. By doing so, we mix players together and thus we are losing individual time-series, but it gives us a good way to see the main path followed by players.

Moreover, such plots are a good way to visually detect outliers. For instance, yellow squares at the top left of fig.3 show that some users have made hundreds of kills using a bow in less than an hour. Those are outliers that need to be removed before further analysis.

VI. ANALYSIS

A. Retention and Quit Rate

We see in fig. 4 that Far Cry 4 quit rate is mainly composed of two peaks : one between 1 and 10 hours of playtime, and one between 30 and 60 hours of playtime. The second peak corresponds to players having completed the game. However, game completion is a subjective notion. Different players need different playtime to complete the game not only because some of them reach a specific goal faster than the others but also because players have different play styles, and thus different goals [2]. Some players, mostly interested int the shooting gameplay, might spend a long time to complete all the side quests as long as they involve shooting on NPCs, while others may be mainly motivated by the main storyline and lose interest as soon as the main quest is complete.

The first peak between 1 and 10 hours corresponds to players who stop the game earlier. From such a high level
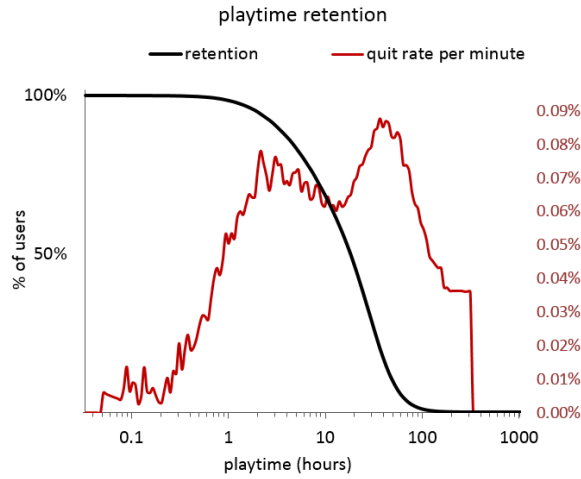
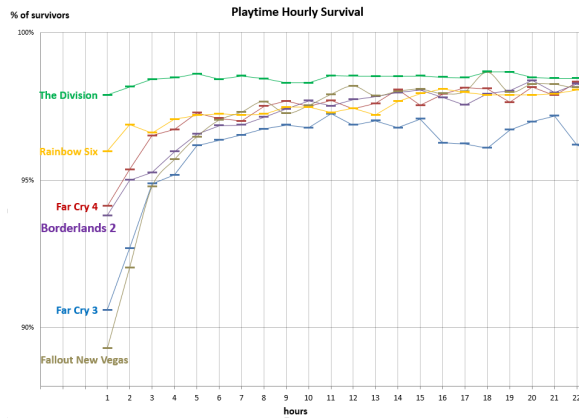Fig. 4. Far Cry 4 retention in black and quit rate in red.

Fig. 5. Hourly survival curve of Far Cry 4 and some related games using a linear scale.

point of view, it is very hard to know why these players quit the game between 1 and 10 hours. But as we know that the quit rate is quite high it might be interesting to investigate more thoroughly in these time ranges, and realize a more detailed survival analysis.

B. Benchmark

We realized a benchmark using SteamSpy data and compared the hourly survival curves as can be seen in (fig.5)

Far Cry 4 has a better hourly survival than Far Cry 3. That's true in the long term as in the first hours of the game. The biggest progress have been made in the first two hours of the game.

Borderlands 2 has quite similar hourly survival that Far Cry 4, but more players quit the game between two and five hours.

Tom Clancy's Rainbow Six® : Siege is also a first person shooter but the gameplay is more multiplayer oriented than Far Cry series. Even though Tom Clancy's Rainbow Six : Siege has almost the same long term hourly survival as Far Cry 4, we can see that Rainbow Six has a better retention in the first two hours than Far Cry 4.

Tom Clancy's The Division is an open world third-person shooter video game, who also includes action-RPG elements. It globally has a much higher retention than Far Cry 4.

We do not see any particular inflexion in Far Cry 4 curves. All the curves globally share a similar shape, and the game is globally among the other similar titles released by Ubisoft. As a result, it is very hard to derive a specific design insight from this high level point of view. In the next section, we will thus focus on a more detailed analysis using survival analysis.

C. Time varying Cox model

We run our model on Weapon-related metrics. Estimated coefficients are plot in fig. 6 and 7. Recall that a positive coefficient means a positive correlation with player departure. As we prefer to talk about variable effect on retention, we should take the coefficient opposite value.

Three weapon usages have a highly positive link with player retention : Machete, Rocket Launcher and Shotgun. The user can only use the machete on the arena so this weapon usage gives us another information about a specific aspect of the gameplay. Playing arena mode allows the player to unlock some exclusive weapons. Players who invest time to unlock new weapons are likely to continue playing for a while.

Rocket and shotgun usage can be related to rough play. It seems that it is a well appreciated play style since players who do many kills with those weapons tend to play longer than others. On the other hand, some weapons related to stealth gameplay, like silencer and sniper rifle, tend to have no positive link on player survival.

Few covariates are still significant on long term - after 20 hours of playtime - except the number of kills made by grenade or machete. Part of this effect is due to the fact that as playtime increases the number of remaining player in the game decreases and hence confidence interval increases.

Auto crossbow is the only one that has a negative link with retention. Note that it has no effect before 10 hours of playtime because it is in average the time needed by players to unlock the weapon. The reason of this negative link could be due to multiple factors. It could thus interesting to perform a playtest in order to find why players who kill lots of enemies using auto crossbow are more likely to stop playing than others.

VII. Conclusion

We've proposed a method to analyze the influence of design and players behavior on retention. First, we have defined player retention and have shown how we could use it to derive quit rate and hourly survival. Quit rate is very useful to detect design issues that are well localized in time, as a peak in the curve describes a sudden loss of players. Hourly survival is more adapted to data set where playtime measures are less accurate, as it is the case, for instance, with data extracted from SteamSpy website. Second, we have proposed a model based on survival analysis to evaluate the link between game variables and player retention, that extends Cox model to deal with time varying game variables and huge datasets.
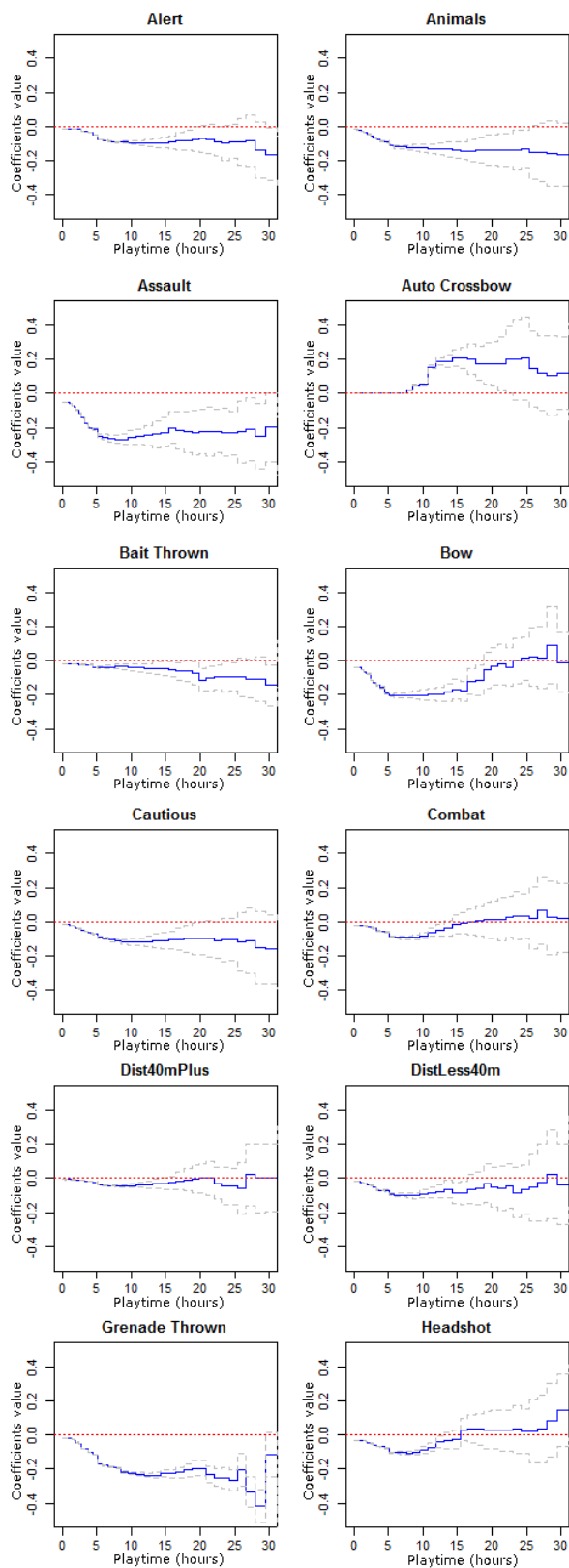
Fig. 6. Time varying coefficients (blue line) estimated on Far Cry 4 weapon usage dataset. Dashed grey line represent 95 % confidence intervals. Red dotted line at zero is the reference for null effect.
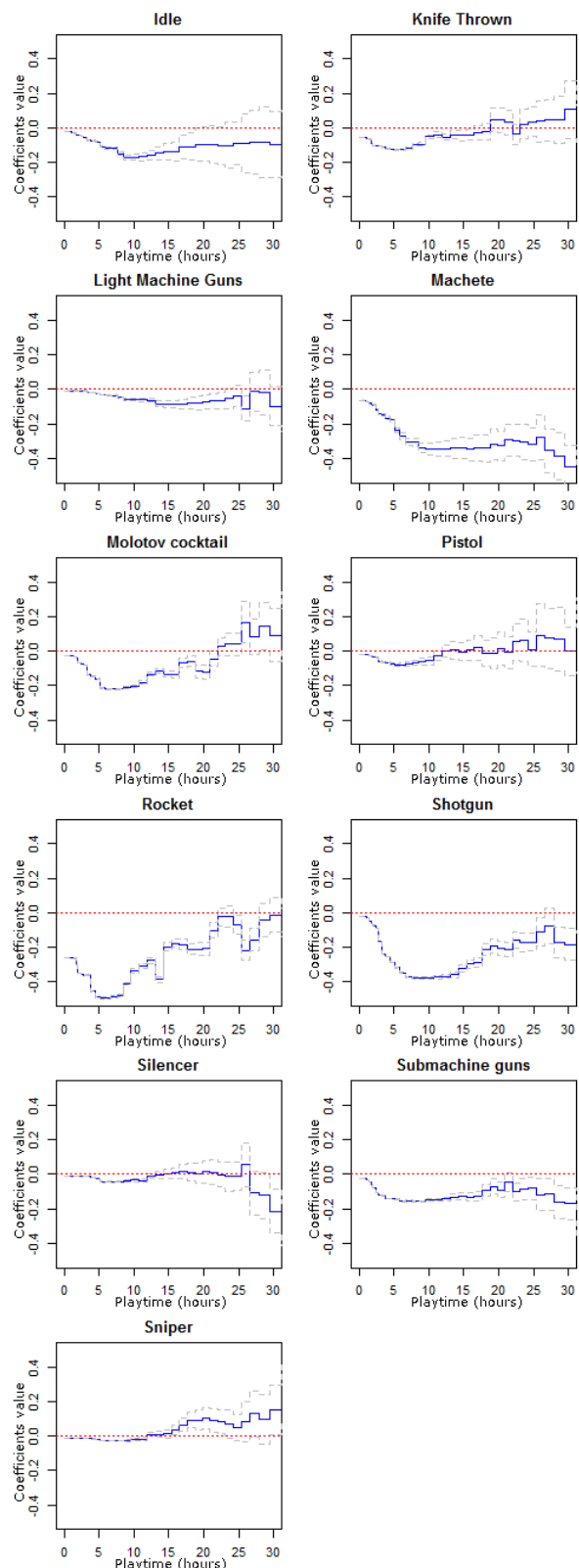

Fig. 7. Time varying coefficients (blue line) estimated on Far Cry 4 weapon usage dataset. Dashed grey line represent 95 % confidence intervals. Red dotted line at zero is the reference for null effect.

Then we used the proposed method to analyze Far Cry 4. First, we analyzed the Quit Rate curve, and found out that there was a first quit rate peak in the early phases of the game, between 1 and 10 hours of play, which might be worth of a more thorough investigation from the development team. Then,the Hourly Survival curve did not reveal any specific issue. This curve is the most easy to compute but also the most coarse grained metric.

We then investigated the game more thoroughly with regard to weapon related game variables. We selected these variable as they concern Far Cry 4 core gameplay. From these data, we found out that many weapons had a positive link with player retention, and only one weapon were found as having a negative link with retention. Globally, it seems that weapons related with stealth gameplay tend to have a less positive link with retention.

We've shown that our approach works well with low level data, like weapon usage. But to get more influent design feedback, we need to introduce higher level metrics.

Difficulty is one of the main factor of motivation [13]. We plan to investigate whether an objective measure of difficulty, as the one proposed by Levieux [11], might have a strong link with player retention.

Another research area that might be very interesting to investigate would be to transform a survey-based theory like Self-Determination Theory [16] into gameplay metrics. Having metrics related to autonomy, competence and relatedness might give interesting insight on player behaviors.

REFERENCES

[1] M. Z. Alaya, T. Allart, A. Guilloux, and S. Lemler. Time-varying high-dimensional aalen and cox models with change-points. *preprint*, 2016.

[2] Richard Bartle. Hearts, clubs, diamonds, spades: Players who suit muds. *Journal of MUD research*, 1(1):19, 1996.

[3] Christian Bauckhage, Kristian Kersting, Rafet Sifa, Christian Thurau, Anders Drachen, and Alessandro Canossa. How players lose interest in playing a game: An empirical study based on distributions of total playing times. In *Computational Intelligence and Games (CIG), 2012 IEEE conference on*, pages 139–146. IEEE, 2012.

[4] Brian Bowman, Niklas Elmqvist, and TJ Jankun-Kelly. Toward visualization for games: Theory, design space, and patterns. *IEEE transactions on visualization and computer graphics*, 18(11):1956–1968, 2012.

[5] Kuan-Ta Chen, Polly Huang, and Chin-Laung Lei. Effect of network quality on player departure behavior in online games. *Parallel and Distributed Systems, IEEE Transactions on*, 20(5):593–606, 2009.

[6] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):pp. 187–220, 1972.

[7] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 12:2121–2159, July 2011.

[8] Dirk Eddelbuettel and Romain François. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18, 2011.

[9] Sergey Galyonkin. Steamspy, 2016.

[10] Brent Edward Harrison and David Roberts. Analytics-driven dynamic game adaption for player retention in a 2-dimensional adventure game. In *Tenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2014.

[11] Guillaume Levieux. *Mesure de la difficulte des jeux video*. PhD thesis, Conservatoire national des arts et metiers-CNAM, 2011.

[12] Tobias Mahlmann, Anders Drachen, Julian Togelius, Alessandro Canossa, and Georgios N Yannakakis. Predicting player behavior in tomb raider: Underworld. In *Computational Intelligence and Games (CIG), 2010 IEEE Symposium on*, pages 178–185. IEEE, 2010.

[13] Thomas W Malone. Toward a theory of intrinsically motivating instruction*. *Cognitive science*, 5(4):333–369, 1981.

[14] T. Martinussen, T. H. Scheike, Ib M. Skovgaard, and T. Matinerssen. Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models. *Scandinavian Journal of Statistics*, 29(1):pp. 57–74, 2002.

[15] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[16] Richard M Ryan and Edward L Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1):68, 2000.

[17] Katie Salen and Eric Zimmerman. *Rules of Play : Game Design Fundamentals*. The MIT Press, October 2003.

[18] G Wallner and S Kriglstein. Visualization-based analysis of gameplay data–a review of literature. *Entertainment Computing*, 4(3):143–155, 2013.

[19] Ben George Weber, Michael John, Michael Mateas, and Arnav Jhala. Modeling player retention in madden nfl 11. In *IAAI*, 2011.

[20] Hanting Xie, Sam Devlin, Daniel Kudenko, and Peter Cowling. Predicting player disengagement and first purchase with event-frequency based data representation. In *Computational Intelligence and Games (CIG), 2015 IEEE Conference on*, pages 230–237. IEEE, 2015.

[21] Nick Yee. Motivations for play in online games. *CyberPsychology & behavior*, 9(6):772–775, 2006.

[22] M. D. Zeiler. Adadelta: An adaptive learning rate method. *CoRR*, abs/1212.5701, 2012.