Bus Inf Syst Eng 58(2):107–118 (2016)
DOI 10.1007/s12599-015-0395-z

CrossMark

RESEARCH PAPER

Making Digital Freemium Business Models a Success: Predicting Customers' Lifetime Value via Initial Purchase Information

Sebastian Voigt · Oliver Hinz

Received: 13 January 2014 / Accepted: 12 April 2015 / Published online: 28 July 2015

Abstract In digital freemium business models such as those of online games or social apps, a large share of overall revenue derives from a small portion of the user base. Companies operating in these and similar businesses are increasingly constructing forecasting models with which to identify potential heavy users as early as possible and create special retention measures to suit those users' needs. In our study, we observe three digital freemium companies that sell virtual credits and investigate to what extent initial purchase information can be used to determine a given customer's lifetime value. We find that customers represent higher future lifetime values if they (a) make a purchase early after registration, (b) spend a significant amount on their initial purchase, and (c) use credit cards to purchase credits. In addition, we see that users tend to spend increasing amounts on subsequent purchases.

Keywords Freemium · Digital business models · Customer lifetime value · Forecasting model

Accepted after two revisions by Prof. Dr. Bichler.

Electronic supplementary material The online version of this article (doi:10.1007/s12599-015-0395-z) contains supplementary material, which is available to authorized users.

Dipl.-Wirtsch.-Inform. S. Voigt (✉) · Prof. Dr. O. Hinz
Chair of Information Systems, Electronic Markets, Technische Universität Darmstadt, Hochschulstraße 1, 64289 Darmstadt, Germany
e-mail: voigt@emarkets.tu-darmstadt.de

Prof. Dr. O. Hinz
e-mail: hinz@emarkets.tu-darmstadt.de
URL: http://www.emarkets.tu-darmstadt.de/

1 Introduction

The growth of the Internet presents a myriad of opportunities for digital business models, along with intensified competition and an accelerated pace of technological change (Veit et al. 2014). These digital business models are highly relevant to IS and BISE researchers as they differ fundamentally from conventional ones, and are particularly challenging with regard to business operations and revenue generation (Amit and Zott 2001; Teece 2010; Veit et al. 2014). In this context, the freemium pricing model (also known as 'free to play' in the gaming industry) has gained increasing attention.

In recent years, freemium has become the dominant pricing strategy for software, games, and social network apps (Guo and Barnes 2009; Lehdonvirta 2009; Lehmann and Buxmann 2009). Users can access such services for free, and usually use them as long as they like at no cost. Wagner et al. (2014) describe the freemium business model as a promising solution for content providers to earn money from users who have a 'for free' mentality. It allows companies to suppress entry barriers and thus to attract a much larger audience than paid services do. However, only a minority of such users actually become paying customers. Among these, we find a large heterogeneity in terms of total revenue spent. The earliest possible identification of high-value customers is a key customer relationship management (CRM) challenge (Goldstein and Gigerenzer 2009). It is a prerequisite for differentiating more profitable from less profitable customers (Kim et al. 2006; Chan and Ip 2011) and for treating the 'best' customers in such a way that they continue paying for the service (Reinartz and Kumar 2003; Malthouse and Blattberg 2005).

Customer relationship management has emerged as an important field at the intersection of marketing and BISE

(Gneiser 2010). Within CRM, customer lifetime value (CLV) is used as a key metric to assess the return on investment of management decisions (Gupta et al. 2006; Gneiser 2010) and marketing measures (Rosset et al. 2003; Zhang et al. 2010), and to optimize long-term customer profitability (Rust et al. 2010; Chan and Ip 2011). CLV represents a customer's present value in terms of expected benefits less the burdens and plays a key role in customer acquisition decisions (Dwyer 1997). (Freemium) companies that can accurately predict if a given new customer will generate future revenue or not, will be able to optimize profitability (Jain and Singh 2002; Malthouse and Blattberg 2005). Therefore, Shaw et al. (2001) propose that companies should develop IS-based support systems that create customer profiles and compute CLVs in order to make better marketing decisions, such as developing VIP programs for promising customers (Zhang et al. 2010) or setting up differentiated direct marketing campaigns.

Especially for digital companies utilizing a freemium model, it is increasingly important to forecast the revenue they can expect from new users which should at least cover acquisition costs (Zhang et al. 2010). Several online blogs have recently reported a steep increase in user acquisition costs. For example, Hochman (2015) mentions an average CPC (cost per click) increase from \$0.38 in 2005 to \$0.92 in 2013 (i.e., an average annual increase of 11.7 %). Likewise, search engine marketers estimate an annual 'Google CPC inflation' of 5–12 % per year (Pretorius 2013). Companies without proper CLV-predicting models risk acquiring unprofitable customers.

Developing forecasting models has been subject of extensive recent research within the IS and BISE community (e.g., Ngai et al. 2009; Lessmann and Voß 2010; Chan and Ip 2011; Gerlach et al. 2013; Cleophas and Ehmke 2014). To predict a customer's future value, digital companies can use various sources of information, among them demographic information, the acquisition channel, or stated preferences. The most frequently used input for CLV models is information on customers' past purchases (Rosset et al. 2003; Malthouse and Blattberg 2005; Wübben and von Wangenheim 2008; Zhang et al. 2010; Chan and Ip 2011). Kim et al. (2006) state that using a CLV calculation based on socio-demographic information and purchase history is more meaningful for segmenting the customer base than all other mentioned methods. Schmittlein and Peterson (1994) point out that past purchase behavior generally outpredicts geo-demographic information.

Although the body of research on CLV is extensive and reached its peak in the 2000s (e.g., Reinartz and Kumar 2000, 2003; Malthouse and Blattberg 2005; Gupta et al. 2006), empirical studies examining Internet-specific business models are still very scarce. In our study, we aim to assess the extent to which a customer's initial purchase information can be used to predict his/her future CLV in a freemium business model, and to study how paying users' purchase amounts evolve with subsequent transactions. To do so, we use three kinds of regression models (NB, OLS, and Logit) and apply them to payment data collected from three European digital service providers (two gaming companies and a dating platform) operating freemium business models. Combined, they have more than 1.3 million registered users, approx. 57,000 of whom are paying customers, spending more than €3 million in more than 195,000 credit purchases. Investigating varied digital businesses allows us to generalize our results.

The remainder of this paper is organized as follows. In Sect. 2, we propose an analytical CLV model specific to digital freemium business models. Section 3 provides a description of the three data sets analyzed for the purposes of our study. Section 4 delineates the different regression models used for empirical analysis. The results of said analysis are addressed in Sect. 5, including CLV prediction and purchase amount development over time. In Sect. 6, we discuss the implications of our results for theory and practice; we also propose future research topics. Finally, the paper closes with a summary.

2 Analytical Model Depicting CLVs for Digital Freemium Business Models

Customer lifetime value can be calculated in numerous ways, depending on the relevant company's industry and price model (Berger and Nasr 1998; Gupta et al. 2006). Jain and Singh (2002) propose the following basic CLV model, which includes the absolute sum of all customer-related, discounted cash flows:

$$\mathrm{CLV} = \sum_{t=1}^{T} \frac{(\mathrm{Rev_t} - \mathrm{C_t})}{(1+d)^{t-0.5}} \quad (1)$$

with $t =$ the period of cash flow (with $t = 1$ as the day of the initial purchase); $T =$ the total number of periods of projected life for the customer under consideration; $Rev_t =$ the revenue from the customer in period t; $C_t =$ the marginal cost of generating the revenue $Re\ v_t$ in period t; $d =$ the discount rate.

Literature distinguishes between CLV definitions for contractual and non-contractual products or services. Contractual products possess continuous cash flow – for instance, insurance policies, mortgages, and cell phone contracts (Rosset et al. 2003; Venkatesan and Kumar 2004; Fader and Hardie 2007). Non-contractual customer relationships, like those of freemium businesses selling virtual credits, are not governed by a membership or contract (Reinartz and Kumar 2003). As such, companies do not observe an explicit customer defection. Purchase timing and amount are not continuous or transparent, and can only be predicted probabilistically (Reinartz and Kumar 2000, 2003; Fader et al. 2005; Borle et al. 2008).

Based on the CLV model discussed above, we propose an adapted model that fits the non-contractual freemium context. First, we refrain from including marginal costs in the CLV equation. For digital businesses, fixed overhead costs such as development, server maintenance, and personnel are more relevant than the marginal costs of reproducing and distributing services over the Internet (Lehmann and Buxmann 2009). We therefore set marginal costs at 0 and focus our CLV assessment purely on revenue, as many analytical CLV models do (Borle et al. 2008).

Second, we need to consider that in a freemium business model, a customer can make multiple purchases on the same day. To calculate the expected future CLV at a specific point in time (e.g., at the initial purchase), it does not suffice to look only at a day-by-day basis, so we include the purchase instance $pi = \{1,\ldots, PN\}$ as a variable to facilitate the correct ordering of intra-day purchases. Furthermore, the relationship does not start with the customer's first purchase, as in retail businesses, but with his/her registration to the (free) service, and ends with his/her final login.

Third, customer lifetime cycles vary depending on the nature of the business and the profile of its customers (Jain and Singh 2002). Naturally, discount rates are considered in businesses with long-lasting customer relationships. Older studies (Kim et al. 1995; Berger and Nasr 1998; Reinartz and Kumar 2000, 2003) use annual discount rates in the range of 12–20 %. In online games and dating, the average user does not stay with the service for very long. In our data sets, between 40 and 55 % of newly registered users log in only once and then never return, and only 1–2 % of all users still generate revenue more than a year after their registration. Blattberg et al. (2009) state that discount rates do not make "much of a difference" for such short time spans, especially given today's low interest rates. Accordingly, we use no discount rates in our CLV model.

With these three adaptations to the original CLV model, the CLV formula for non-contractual freemium business models is as follows:

$$\mathrm{CLV}_{\mathrm{freemium}} = \left(\sum_{t=1}^{T} \sum_{\mathrm{pi}=1}^{PN} \mathrm{Re}v_{t,\mathrm{pn}} \right) \tag{2}$$

with $t =$ the period of cash flow (with $t = 1$ as the day of customer's registration); $T =$ the total customer lifetime of the customer under consideration; $\mathrm{pi} =$ the customer's purchase instance; $\mathrm{PN} =$ the total number of purchases the customer makes within T; $\mathrm{Re}\ v_t =$ the revenue from purchase pi in period t.

3 Data Description

For our empirical study, we use customer and purchase data from three European digital businesses that sell a virtual currency in a non-contractual freemium model. A data confidentiality agreement with the companies prevents us from disclosing their names. Working with multiple data sets allows us to see whether our results can be applied broadly rather than being specific to one platform. The three data sets come from two online real-time strategy games (data sets 1 and 2), and an online dating platform (data set 3).

The services can be used via Internet browsers and dedicated mobile apps. They feature a shop where users can buy credits that provide them with access to additional in-app features compared to non-paying users, such as extra resources for constructing cities or armies more quickly (gaming), or the ability to send messages to other users (dating).

The selection of the gaming and dating software industries for our study covers the most in-demand digital service applications and provides us with highly relevant data. A look at the top-grossing apps for iPhone (AppAnnie 2015) in the US, Japan and Germany shows that games constitute by far the most popular apps (in terms of generated revenue), followed by social apps (i.e., networking and dating). We find in Table 1 that on average, 94 % of the top-grossing apps are free to download, leaving only 6 % that require an upfront payment. 97 % of all apps allow in-app purchases (IAP).

When registering for one of the two online games in the study, users must provide no more than a nickname and an email address. Additional demographics are only very rarely revealed; as a result, the companies' knowledge of their user base is very limited. On the dating platform (data set 3), users usually share profile pictures, gender, age, place of residence, and other personal details; however, this

Table 1 Top-grossing iPhone apps in Apple's AppStore – 4 January 2015

Country	Genre	Free	Paid	Have IAP
US	Gaming	82	4	83
Top 100	Social/dating	8	1	9
	Other	4	1	4
Japan	Gaming	85	0	85
Top 100	Social/dating	14	0	14
	Other	1	0	1
Germany	Gaming	68	3	69
Top 100	Social/dating	8	3	10
	Other	12	6	16
Share of all apps		94 %	6 %	97 %

information is – according to the platform provider – inaccurate or incomplete for the majority of the new registrations.

When buying credits on any of these platforms, users choose a purchase amount (e.g., €5, €10, €20) and a payment method (e.g., credit card, PayPal, prepaid card). The available amounts differ according to the payment method. Payment methods are available for data sets 1 and 2, but not for data set 3. We excluded all incomplete purchases, such as chargebacks and free promotions (with a revenue of €0) from our samples.

For each user we consider all purchases he/she makes within the first 365 days after his/her registration as CLV. This 1-year window is in line with previous CLV studies (e.g., Najar and Rajan 2005; Rust and Verhoef 2005). For data set 3, which is the smallest, we reduce the CLV calculation window to 180 days in favor of expanding the observation timeframe to compensate for the disparity in user count. On average, the paying users in data set 3 make their last purchase 39.3 days after registration (SD 48.3 days), so only negligible revenue is lost through our adjustment.

Table 2 shows that in our data sets, only 3.3–6.1 % of registering users become paying users (that is, purchase credits at least once). A business with >90 % free users sharing little to no demographic information makes it difficult to predict a user's future value to the company. Since the goal of our paper is to predict CLV using paying customers' initial purchase information, we exclude free users (whose CLV is always 0) from our data set for all analyses (same approach as Fader et al. 2005).

All three data sets show significant heterogeneity among paying users. In the paying user base, average CLV varies from €31.01 (data set 2) to €55.25 (data set 3), with standard deviations between €52.08 (data set 2) and €124.39 (data set 1). In data set 1, 20 % of the paying customers with the highest CLV contribute up to 75.6 % of total revenue. When we consider all users (including non-paying), we see that the top 1 % of users contribute up to 84.6 % (data set 1), 60.5 % (data 2), and 53.6 % (data set 3) of total revenue. The numbers indicate how important it is for freemium companies to identify high-value customers expediently in order to design proper CRM measures that suit varied customer segments.

Despite operating similar business models, freemium gaming and dating platforms are quite distinct from a user perspective. For example, whereas a 'successful' gamer is likely to keep playing (Choi and Kim 2004), a 'successful' dater is likely to quit the service after finding a partner. It will be interesting to see if these conflicting user objectives affect CLV or if we find consistent results for both businesses.

4 Methodology and Empirical Model Description

We estimate a negative binomial (NB) regression model to assess how past purchase information impacts a paying user's future CLV at the time of his/her initial credit purchase (i.e., pi = 1 in our CLV model). As a user's demographic information is often incomplete or incorrect for non-contractual online services, we focus purely on past purchase behavior to predict users' future CLV. Following our previous argumentation, we use three input variables:

1. *Time until initial purchase* (*in days*): Equals a user's past customer lifetime (in days) elapsed before he/she makes his/her initial purchase.
2. *Initial purchase amount* (*in* €): Equals the chosen virtual currency package (e.g., €10, €20, and €50) for a user's initial purchase.
3. *Payment method used for initial purchase*: Equals the chosen payment method (see Table 1 in the Online Appendix for a detailed description) when purchasing credits for the first time. The payment methods are coded as dummy variables. Each is set at 1 if used, and 0 if not.

Wübben and von Wangenheim (2008) describe NB models as the state-of-the-art approach in determining the future activity and purchase levels of a customer. Based on the work of Schmittlein et al. (1987) and Schmittlein and Peterson (1994), these kinds of models have been employed in several studies (Reinartz and Kumar 2000, 2003; Fader et al. 2005; Glady et al. 2009; Zhang et al. 2010). NB models are attractive because they (1) forecast individuals' future purchase levels and (2) operate on past transaction behavior. They operate solely on the frequency and recency of a customer's past purchase behavior (Wübben and von Wangenheim 2008), and are thus well-suited to forecasting CLV in a freemium business model.

Gupta et al. (2006) summarize frequently-used alternative modeling approaches that aim to predict CLV, among them probability, econometric, persistence, computer science, and diffusion/growth models. Goldstein and Gigerenzer (2009) name studies (such as Venkatesan and Kumar 2004) in which linear regression models fit to predict future 'best' customers and their purchase activity. Jain and Singh (2002) recommend such regression models and Bayesian approaches (as applied by Borle et al. 2008) to assess which individuals in the customer group most likely represent active and inactive customers, and what level of transactions the company can expect from them in the future, both individually and collectively.

In all of our data sets, the standard deviation exceeds the mean in terms of CLV per paying user (see Table 2 above). Such customer heterogeneity is accepted when predicting CLV (Reinartz and Kumar 2003). NB models possess an

Table 2 Descriptives

	Data set 1	Data set 2	Data set 3
Observation timeframe (for users to register)	2009-01-01 to 2010-10-31	2009-01-01 to 2010-10-31	2011-09-25 to 2012-07-04
Newly registered users within timeframe	831,617	444,934	52,200
Timeframe for CLV calculation (after registration)	365 days	365 days	180 days
Share of newly registered users becoming paying users	3.3 %	6.1 %	4.8 %
Sample size			
Credit purchases	104,659	87,230	3,639
Paying users	27,707	27,236	2,500
Revenue	€1.3 m	€845 k	€138 k
CLV per paying user			
Min	€2.99	€2.99	€9.99
Max	€10.932,76	€1,102.29	€1,697.00
Mean	€47.01	€31.01	€55.25
Median	€10.97	€10.98	€29.00
SD	€124.39	€52.08	€73.20
Credit purchases per paying user			
Min	1	1	1
Max	224	96	23
Mean	3.78	3.20	1.46
Median	2	2	1
SD	6.84	4.58	1.29
Avg. purchase amount			
All purchases	€12.45	€9.68	€37.96
Initial purchases	€9.46	€9.08	€34.81
Non-initial purchases	€13.52	€9.96	€44.87

extra parameter to accommodate such overdispersion (STATA 2015; UCLA 2015) and are thus especially useful for discrete, overdispersed data, where a Poisson distribution is unsuitable. This also allows us to use discrete data as dependent variables, such as purchase amounts in € which depend on the chosen payment method.

In the second part of our empirical work (in Sect. 5.2), we assess if and how a user's purchase amounts change over time, in the case that he/she makes multiple credit purchases. We apply a standard OLS regression with *Purchase amount* of a given transaction (in €) as dependent variable, while a user's *Purchase instance* (i.e., whether it was his/her first, second etc. transaction) and *Payment method used for initial purchase* serve as predictive variables.

Finally, we apply a logistic regression to assess the probability of a customer's purchase amount changing over time. Logit models are non-linear estimation techniques with a binary outcome (Zhang et al. 2010), which solve the problem of unboundedness of OLS. Alternatively, a Probit model could be used, but both models usually yield very similar results (Freedman 2009, pp. 121–129). The dependent dummy variable *Same amount as previous purchase* equals 1 if a purchase amount is identical to the previous purchase, and 0 if it is higher or lower. The *Payment method used* (also coded as a dummy variable) for the purchase in question is included as a control variable.

To address the challenges that outliers pose for some statistical models, we use the Huber-White sandwich estimators (Huber 1967; White 1980) in all three regression models, thereby obviating minor concerns about the potential failure to meet assumptions, such as normality, heteroskedasticity, or observations that exhibit large residuals, leverage, or influence. For all regression models and data sets, we check the correlation matrices to identify potential multicollinearity issues. For all (non-dummy) independent variables, the off-diagonal correlation values are all clearly below the common threshold of 0.4, which would indicate multicollinearity problems (Fickel 2001, p. 41).

5 Empirical Results

5.1 CLV Prediction Using Initial Purchase Information

For our first analysis, we check the influence of the information available at the initial credit purchase upon a

Table 3 Results from NB regression (Dep. var.: *CLV in €*)

	Data set 1	Data set 2	Data set 3
Time until initial purchase (in days)	0.956***	0.960***	0.8550***
Initial purchase amount (in €)	1.0354***	1.0269***	1.0119***
Payment method used for initial purchase			
Bank transfer	0.017	0.109***	
Credit card	1.2206***	1.0487	
Direct debit	0.987	0.943***	
Phone	0.013***	0.781***	
Cell phone	0.987*	0.552	
Prepaid card	0.478	0.974***	
SMS	0.522***	0.098	
Wallet	(omitted)	(omitted)	
Constant	40.46722***	24.2738***	17.3975***

* $p < .1$; ** $p < .05$; *** $p < .01$

customer's (remaining) CLV. We estimate several NB models to ensure the robustness of our results. We first consider only *Initial purchase amount* as independent variable and successively include *Time until initial purchase* and *Payment method used for the initial purchase*, which results in our final model shown in Table 3. We detect no change of algebraic signs at the significant variables from one model to another and thus believe our model to be robust. The reported incidence rate ratios (IRRs) are the exponential equivalent to regular regression coefficients.

Investigating the *Time until initial purchase (in days)*, we see that the IRRs are less than 1 in all three data sets. This means that the earlier a new user buys credits for the first time (after registration), the higher his/her remaining CLV. Conversely, this means that users who decide to purchase credits later than others usually have a lower future CLV.

Result 1 The length of time between a user's registration and his/her initial credit purchase is negatively correlated to that user's remaining CLV.

Moreover, we can see in three data sets (with each $p < .01$) that the *Initial purchase amount* is positively correlated to remaining CLV. The effect is more pronounced in the two online games than in the dating platform. These results show that even in settings with usually short customer relationships, customers spending large amounts in their initial purchase promise higher future revenues than those spending less.

Result 2 The higher the customer's initial credit purchase, the higher his/her remaining CLV.

Data sets 1 and 2 offer eight payment methods each. We can see that in both cases, using a credit card for the first purchase has the highest positive impact on the remaining CLV (data set 1: 1.2206, $p < .01$; data set 2: 1.0487, n.s.), followed by online wallets (omitted variable with an IRR set to 1). SMS payments promise the lowest remaining CLV (data set 1: 0.5522, $p < .01$; data set 2: 0.7098, n.s.). Overall, four out of seven payment methods show significant results in each of the data sets (excl. the omitted Wallet method).

Result 3 The payment method of the initial purchase is a significant predictor of remaining CLV. Credit card users promise higher CLVs than all other paying customers.

5.2 Development of Purchase Amounts in Subsequent Purchases

We now assess to what extent a user's purchase amounts develop over subsequent transactions. The last row in Table 2 shows that the average amount (in €) of all users' initial purchases is lower than the amount of each subsequent purchase. On average, the amount of any subsequent purchase is 42.9 % (data set 1), 9.7 % (data set 2), and 28.9 % (data set 3) larger than the initial purchase.

Figure 1 illustrates how average purchase amounts develop for all three data sets. We indexed the average amount spent on the initial purchase at 1 (e.g., €9.46 for data set 1). For all platforms, we can see that the average purchase amount increases with the second and third purchases, and more or less flattens after that. For example, the average amount spent in the tenth purchase is 57 % higher for data set 1, 5 % higher for data set 2, and 41 % higher for data set 3 than in the initial purchase. The results of the OLS regression in Table 4 confirm an increase of the average *Purchase amount in €* with subsequent purchases.

Result 4 The average revenue per purchase increases with the number of purchases a customer makes.

Finally, we aim to assess the probability of a change in purchase amount for a customer's subsequent purchases. We first compare how the purchase amount develops from one purchase to another. If a customer decides to make a purchase beyond the first one, he/she has three options: (1)

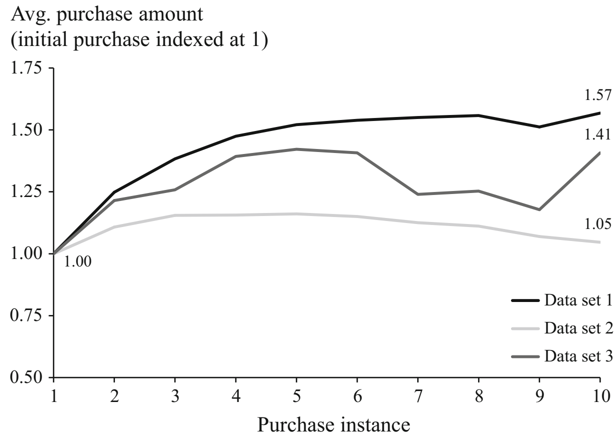

Fig. 1 Average purchase amount depending on purchase instance

purchase credits for the same amount as in the previous purchase, (2) spend more than previously, or (3) spend less. Figure 2 illustrates how often each of the cases occurs in the three data sets. In data set 1, for example, we see that from all customers making a second purchase, 59.6 % spent the same amount in the second purchase as in the first, 28.1 % spent more, and 12.3 % spent less. We find a consistent picture across all three data sets: the more purchases a customer has made, the less likely he/she is to spend more than in his/her previous purchase. However, this is still more likely than a drop in the purchase amount.

To check if the descriptive results from Fig. 2 are statistically significant, we conduct a logistic regression. In Table 5, we can see that *Purchase instance* has a positive and highly significant impact on the probability that the per-instance purchase amount remains stable. The likelihood of a purchase amount change reduces with every purchase. For data set 3, the coefficient is also positive but the results are not statistically significant. Changing the dependent variable into *Higher/lower amount than previous purchase*, we see that the *Purchase instance* has a negative impact on both the likelihood of the user choosing a higher (data sets 1 and 2: $p < .01$; data set 3: $p < .1$) or a lower purchase amount (data set 1: $p < .01$; data sets 2 and 3: n.s.).

Result 5 With each subsequent purchase, the purchase amount becomes less likely to change (compared to the previous purchase).

6 Discussion

6.1 Theoretical Contributions

Our empirical analysis produces five key results, which are consistent for two quite distinct businesses: online gaming and dating (except result 3, for which data for the dating platform is unavailable). We thus believe that the results may also apply to other freemium business models.

First, the sooner a user becomes a payer, the higher his/her remaining CLV. This result agrees with existing theory from traditional businesses which incur an upfront payment. Early adopters of a product or service also tend to be heavy users (Taylor 1977) and less price-sensitive than late

Table 4 Results from OLS regression (Dep. var.: *Purchase amount in €*)

	Data set 1	Data set 2	Data set 3
Number of observations	104,659	87,230	3639
Prob > F	0.000	0.000	0.000
R square	0.483	0.443	0.444
Root MSE	11.445	8.6473	23.022
	Coeff.	Coeff.	Coeff.
Purchase instance	0.544***	0.480***	2.4477***
Payment method used for initial purchase			
Bank transfer	(omitted)	(omitted)	
Credit card	4.9880***	0.887	
Direct debit	0.613	−3.4678***	
Phone	−11.9071***	−11.8921***	
Cell phone	−9.8883***	−11.7784***	
Prepaid card	−4.4203***	−6.6812***	
SMS	−14.8287***	−14.9902***	
Wallet	−1.9824***	−4.6504***	
Constant	21.6124***	20.9884***	33.5591***

* $p < .1$; ** $p < .05$; *** $p < .01$

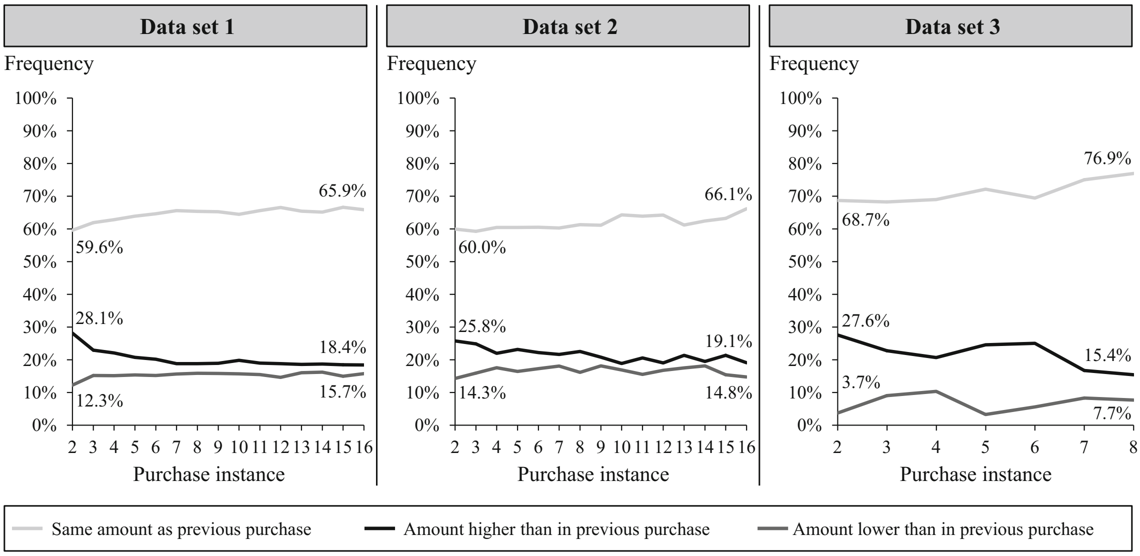

Fig. 2 Probability of a purchase amount change depending on purchase instance

Table 5 Results from logistic regression (Dep. dummy var.: *Same amount as previous purchase*)

	Data set 1	Data set 2	Data set 3
Purchase instance	0.096***	0.068***	0.265
Payment method used			
Bank transfer	−0.383***	−0.393***	
Credit card	−0.275***	−0.972***	
Direct debit	−0.984**	−0.025*	
Phone	−0.316***	−0.065***	
Cell phone	−0.938***	−0.677***	
Prepaid card	−0.102***	−0.215***	
SMS	0.273***	0.989***	
Wallet	(omitted)	(omitted)	
Constant	0.810***	0.486***	0.201***

* $p < .1$; ** $p < .05$; *** $p < .01$

adopters (Goldsmith and Newell 1997). In the case of freemium business models, users who decide to move from being free users to paying customers very quickly can be considered early adopters of the paid part of the service. According to our results, they are likely to take the service more seriously than those who make the transition later, and thus are likely to spend more money in total.

Second, we find that the higher a customer's initial purchase amount is, the higher his/her remaining CLV. That this holds true in all three cases is particularly interesting, as Malthouse and Blattberg (2005) describe a contradictory situation in which customers with exceptional spending in a given instance often converge to a lower true mean spending over the course of future transactions. One can expect a similar effect with freemium services that do not exhibit sufficient incentives and features on which to spend credits. Various studies support our findings, however. Reinartz and Kumar (2003) find positive relationships between previous spending levels and customer lifetime. Venkatesan and Kumar (2004) show that revenue generated by a certain customer in the past is positively correlated to – and thus a good predictor of – his/her future spending. Blattberg et al. (2009) discover that historical profitability is usually closely correlated with future CLV estimates. Our results are remarkable in the context of online dating, as one could have expected that paying users invest their credits and either leave the service happily (having found a partner) or unhappily (having found no partner despite having bought credits). We assume that the experience for paying users is good enough to keep them engaged and induce them to buy credits again.

Third, we find that the payment method of the initial purchase significantly impacts a user's remaining CLV. Although the results are only partially statistically significant, they indicate that credit card users tend to have the highest CLV, followed by users of e-wallets (such as PayPal), direct debit and bank transfer, prepaid cards, and finally cell phone and SMS payment. The fact that credit card users have a higher willingness to pay (as shown by Feinberg 1986; Prelec and Simester 2001; Soman 2001) does not come as a surprise. For digital business models however, surprisingly little research has been conducted to date on the effects of payment method on a consumer's future behavior.

Fourth, we see that average revenue per purchase increases with the number of purchases a customer makes. This comports with the findings of Reichheld (2001) who attributes the acceleration in customers' revenues over time to (among other things) an increased use of the product and a higher willingness to pay. Some studies (e.g., Reinartz and Kumar 2003; Liu 2007) provide support that the amount spent per purchase is positively correlated to customer loyalty. Fader et al. (2005) analyze the customer base of an online music site and discover a positive correlation between the average purchase amount and the number of purchases.

In our empirical study, we examined online gaming and dating platforms. Characteristics of such services' users may have led or supported the results. Many games are competitive, and those who continue playing a specific game for a long time are highly motivated to attain a certain in-game status and/or achieve a high or the highest possible score or rank (Bostan 2009), which positively influences their willingness to pay for credits. In addition, gaming and dating platforms are social networks that exhibit positive network effects which eventually increase users' willingness to pay (Farrell and Saloner 1985; Bapna and Umyarov 2012). The longer a customer uses the service, the more social contacts he/she may get to know, and the stronger the network effects become. This could mean that freemium business models that exert weaker network effects may see less progressive revenue per transaction.

Fifth and last, our data shows that a customer's purchase amount becomes less likely to change with each subsequent purchase. Digital services such as software, gaming, or dating services are experiential goods whose actual value customers can assess only after their usage or purchase (Lehmann and Buxmann 2009), as are credit purchases. Users who decide to make another credit purchase are likely to have been satisfied with their previous buying decision. According to Latham and Locke (1991), consumers who realize that their purchase behavior is instrumental to achieving a positive outcome (e.g., building a stronger base in an online game or getting in touch with other people on a dating website) will be more likely to engage in this behavior, and therefore make repeated purchases out of habit (Aarts et al. 1998). Still, our results are somewhat limited because the companies we investigated offered only a selected range of credit packages; this limits the customer's ability to increase his/her amount per purchase indefinitely.

6.2 Practical Contributions

Our five main findings allow us to derive several recommendations for operators of digital freemium businesses. Before we propose these, we would like to stress that our research only deals with correlations; we are not able to draw explicit conclusions of causality between initial purchase information and future CLV. For example, we see that credit card users have on average higher CLVs than any other users. However, this does not mean that moving all customers to credit card payments is necessarily profitable (see Takac et al. 2011 for a longer discussion on this 'causality challenge').

In our three data sets, we see that up to 84.6 % of the total revenues derive from a mere 1 % of users. This represents a clear call to action for companies operating a freemium model to identify potential heavy spenders early on and to cater to them as best as possible. We find that payment information at the time of the initial purchase can be used to identify high-value users: customers tend to generate higher future CLVs if they (a) become paying customers early after their registration, (b) spend a large amount on their initial purchase, and (c) use specific payment methods (especially credit cards). These findings can be used in an information system that segments the customer base by forecasted CLV (Rust et al. 2010), executes appropriate CRM measures (Chan and Ip 2011), and evaluates these measures' effectiveness in CLV increase (Zhang et al. 2010).

An appropriate CLV prediction can support freemium businesses' acquisition, retention, and monetization strategies. As freemium companies pay customer acquisition costs despite not knowing if they will ever generate revenues with a given user, it is important that they align these costs with their CLV forecasts to assess whether a certain customer group is likely to generate enough future revenues to be profitable (Zhang et al. 2010). Regarding retention, identified high potential-value customers should be treated with special care – for example, with a (VIP-like) retention program with dedicated services (such as guaranteed response time to service tickets or unique platform content) to reduce their churn risk (Rosset et al. 2003). Observing users' previous purchase amounts and CLV potential also allows freemium companies to apply differentiated monetization strategies for up- and cross-selling. This includes tactics to take payers to the next highest spender bracket (e.g., from a €10 credit package to a €20 package). Overall, we find that a thorough CLV forecast can support and inform a large range of CRM applications.

6.3 Limitations

Our study is the first to predict CLV and purchase amount development over time within the context of digital freemium business models, and leaves room for improvement and future research. The initial-purchase information we used should be available to all freemium companies and thus serves as a sensible starting point. A logical extension would be to update the CLV prediction over time, for example, after the second or third purchase (Borle et al. 2008), or to assess CLV right after a user's registration. This, however, requires additional information on the user such as user demographics (e.g., sex, age, place of residence), and acquisition channel (e.g., SEM, online ads, partner websites). Also, on-site activity (e.g., number of logins, time on platform in the last week) would be interesting input data to explore (Alves et al. 2014). Chan and Ip (2011) indicate that today no information system that forecasts CLV takes all key areas into account at the same time. If a system did so, a CLV model could create a clearly better picture as to which users promise the highest future CLV (Glady et al. 2009).

In accordance with our definition of CLV, we looked solely at the revenue a user generates directly during his/her relationship with the company. However, a user may also generate extra value indirectly, for example by referring the service to other users (and thus reducing the acquisition costs for those users; Hinz et al. 2011), or by exerting network effects that positively influence other users' willingness to pay (Farrell and Saloner 1985). An expansion of the model may be helpful to assess the CLV more holistically.

As mentioned above, our results indicate only correlations, as it is hard to draw causal conclusions based solely on transactional data (Takac et al. 2011). For future research, it would be fruitful to conduct A/B field experiments (e.g., discounting credit packages for users who just registered to see if an earlier payer conversion drives CLV, or excluding certain payment methods to see if shifting customers to other methods increases their CLV), which would ultimately allow causal claims to be made.

7 Summary

This study had two main objectives: First, to assess how information that is available to digital freemium companies at a customer's initial purchase impacts his/her future CLV, and second, to measure how paying customers' purchase amounts develop with subsequent purchases. For that purpose, we collected registration and purchase data from three digital companies (two online strategy games and a dating platform) operating a non-contractual freemium business model. Combined, our three data sets comprised more than 1.3 million user registrations, of which about 57,000 became paying customers. These users processed more than 195,000 credit purchases and spent about €3 million on the services in question.

We used three kinds of regression models (NB, OLS, and Logit) for our empirical analysis. Our results show that digital companies operating a freemium business model can expect a higher future CLV from users who…

1. …become paying customers early after their registration,
2. …spend a large amount on their initial purchase, and
3. …use specific payment methods (especially credit cards).

Our results are consistent for the two rather distinct business types (gaming and dating) examined; we therefore believe the findings can be applied to other digital freemium business models as well.

Beyond predicting CLV at an early stage, we saw that users tend to spend an increasing amount on subsequent purchases. However, the likelihood of the purchase amount increasing (or declining) compared to the last purchase decreases with each additional purchase. We believe that many new users are uncertain about the true value of the credits when they buy them for the first time, and that this uncertainty is dissolved when they decide to make another purchase.

Our data has shown that digital freemium businesses operate in an environment with very heterogeneous users: in one case, 1 % of the user base accounted for almost 85 % of total revenues. This makes clear how important it is to identify high-potential customers as soon as possible and to give them preferential treatment. Freemium companies can integrate our CLV prediction model into their CRM system to adapt their customer acquisition, retention and monetization strategies accordingly. Changes in CLV should be carefully monitored and used to improve the effectiveness of these strategies.

References

Aarts H, Verplanken B, Knippenberg A (1998) Predicting behavior from actions in the past: repeated decision making or a matter of habit? J Appl Soc Psychol 28(15):1355–1374

Alves J, Lange S, Riedmiller M (2014) Case study: behavioral prediction of future revenues in freemium games. In: NC² Workshop, 36th Ger Conf Pattern Recognit (GCPR)

Amit R, Zott C (2001) Value creation in e-business. Strat Manag J 22(6–7):493–520

AppAnnie (2015) iOS top app charts, 4 January 2015. http://www.appannie.com/apps/ios/top/?device=iphone. Accessed 05 Jan 2015

Bapna R, Umyarov A (2012) Do your online friends make you pay? A randomized field experiment in an online music social network. http://www.heinz.cmu.edu/download.aspx?id=3660. Accessed 25 Mar 2015

Berger PD, Nasr NI (1998) Customer lifetime value: marketing models and applications. J Interact Mark 12(1):17–30

Blattberg RC, Malthouse EC, Neslin SA (2009) Customer lifetime value: empirical generalizations and some conceptual questions. J Interact Mark 23(2):157–168

Borle S, Singh SS, Jain DC (2008) Customer lifetime value measurement. Manag Sci 54(1):100–112

Bostan B (2009) Player motivations: a psychological perspective. CIE 7(2):1–26

Chan SL, Ip WH (2011) A dynamic decision support system to predict the value of customer for new product development. Decis Support Syst 52(1):178–188

Choi D, Kim J (2004) Why people continue to play online games: in search of critical design factors to increase customer loyalty to online contents. CyberPsychol Behav 7(1):11–24
Cleophas C, Ehmke JF (2014) When are deliveries profitable? Bus Inf Syst Eng 6(3):153–163
Dwyer RF (1997) Customer lifetime valuation to support marketing decision making. J Interact Mark 11(4):6–13
Fader PS, Hardie BG (2007) How to project customer retention. J Interact Mark 21(1):76–90
Fader PS, Hardie BG, Lee KL (2005) RFM and CLV: using iso-value curves for customer base analysis. J Market Res 42(4):415–430
Farrell J, Saloner G (1985) Standardization, compatibility, and innovation. RAND J Econ 16(1):70–83
Feinberg RA (1986) Credit cards as spending facilitating stimuli: a conditioning interpretation. J Consum Res 13(3):348–356
Fickel N (2001) Sequenzialregression: Eine neodeskriptive Lösung des Multikollinearitäts-problems mittels stufenweise bereinigter und synchronisierter Variablen. VWF, Berlin
Freedman DA (2009) Statistical models: theory and practice. Cambridge University Press, New York
Gerlach M, Cleophas C, Kliewer N (2013) Airline codeshare alliances. Bus Inf Syst Eng 5(3):153–163
Glady N, Baesens B, Croux C (2009) A modified Pareto/NBD approach for predicting customer lifetime value. Expert Syst Appl 36(2):2062–2071
Gneiser MS (2010) Value-based CRM – the interaction of the triad of marketing, financial management, and IT. Bus Inf Syst Eng 2(2):95–103
Goldsmith RE, Newell SJ (1997) Innovativeness and price sensitivity: managerial, theoretical and methodological issues. J Prod Brand Manag 6(3):163–174
Goldstein DG, Gigerenzer G (2009) Fast and frugal forecasting. Int J Forecast 25(4):760–772
Guo Y, Barnes S (2009) Virtual item purchase behavior in virtual worlds: an exploratory investigation. Electron Commer Res 9(1–2):77–96
Gupta S, Hanssens D, Hardie B, Kahn W, Kumar V, Lin N, Ravishanker N, Sriram S (2006) Modeling customer lifetime value. J Serv Res 9(2):139–155
Hinz O, Skiera B, Barrot C, Becker JU (2011) Seeding strategies for viral marketing: an empirical comparison. J Market 75(6):55–71
Hochman J (2015) The cost of pay-per-click (PPC) advertising – trends and analysis. https://www.hochmanconsultants.com/articles/je-hochman-benchmark.shtml. Accessed 25 Mar 2015
Huber PJ (1967) The behavior of maximum likelihood estimates under nonstandard conditions. In: 5th Berkeley Symposium Mathematical Statistics Probability, Berkeley, pp 221–233
Jain D, Singh SS (2002) Customer lifetime value research in marketing: a review and future directions. J Interact Mark 16(2):34–45
Kim N, Mahajan V, Srivastava RK (1995) Determining the going market value of a business in an emerging information technology industry: the case of the cellular communications industry. Technol Forecast Soc 49(3):257–279
Kim SY, Jung TS, Suh EH, Hwang HS (2006) Customer segmentation and strategy development based on customer lifetime value: a case study. Expert Syst Appl 31(1):101–107
Latham GP, Locke EA (1991) Self-regulation through goal setting. Organ Behav Hum 50(2):212–247
Lehdonvirta V (2009) Virtual item sales as a revenue model: identifying attributes that drive purchase decisions. Electron Commer Res 9(1–2):97–113
Lehmann S, Buxmann P (2009) Pricing strategies of software vendors. Bus Inf Syst Eng 1(6):452–462
Lessmann S, Voß S (2010) Customer-centric decision support. Bus Inf Syst Eng 2(2):79–93
Liu Y (2007) The long-term impact of loyalty programs on consumer purchase behavior and loyalty. J Mark 71(4):19–35
Malthouse EC, Blattberg RC (2005) Can we predict customer lifetime value? J Interact Mark 19(1):2–16
Najar V, Rajan MV (2005) Measuring customer relationships: the case of the retail banking industry. Manag Sci 51(6):904–919
Ngai EWT, Xiu L, Chau DCK (2009) Application of data mining techniques in customer relationship management: a literature review and classification. Expert Syst Appl 36(2):2592–2602
Prelec D, Simester D (2001) Always leave home without it: a further investigation of the credit-card effect on willingness to pay. Mark Lett 12(1):5–12
Pretorius K (2013) Google adwords CPC inflation. http://www.mediavisioninteracti-ve.com/blog/paid-search/google-adwords-cpc-inflation/. Accessed 25 Mar 2015
Reichheld FF (2001) The loyalty effect: the hidden force behind growth, profits, and lasting value. Harvard Business Press, Boston
Reinartz WJ, Kumar V (2000) On the profitability of long-life customers in a noncontractual setting: an empirical investigation and implications for marketing. J Mark 64(4):17–35
Reinartz WJ, Kumar V (2003) The impact of customer relationship characteristics on profitable lifetime duration. J Mark 67(1):77–99
Rosset S, Neumann E, Eick U, Vatnik N (2003) Customer lifetime models for decision support. Data Min Knowl Disc 7(3):321–339
Rust RT, Verhoef PC (2005) Optimizing the marketing interventions mix in intermediate-term CRM. Mark Sci 24(3):477–489
Rust RT, Moorman C, Bhalla G (2010) Rethinking marketing. Harv Bus Rev 88(1/2):94–101
Schmittlein DC, Peterson RA (1994) Customer base analysis: an industrial purchase process application. Mark Sci 13(1):41–67
Schmittlein DC, Morrison DG, Colombo R (1987) Counting your customers: who are they and what will they do next? Manag Sci 33(1):1–24
Shaw MJ, Subramaniam C, Tan GW, Welge ME (2001) Knowledge management and data mining for marketing. Decis Support Syst 31(1):131–141
Soman D (2001) Effects of payment mechanism on spending behavior: the role of rehearsal and immediacy of payments. J Consum Res 27(4):460–474
STATA (2015) Nbreg – negative binomial regression. www.stata.com/manuals13/rnb-reg.pdf. Accessed 25 Mar 2015
Takac C, Hinz O, Spann M (2011) The social embeddedness of decision making: opportunities and challenges. Elec Mark 21(3):185–195
Taylor JW (1977) A striking characteristic of innovators. J Mark Res 14(1):104–107
Teece DJ (2010) Business models, business strategy and innovation. Long Range Plan 43(2):172–194
UCLA (2015) Stata data analysis examples: negative binomial regression. http://www.ats.ucla.edu/stat/stata/dae/nbreg.htm. Accessed 25 Mar 2015
Veit DE, Clemons E, Benlian A, Buxmann P, Hess T, Spann M, Kundisch D, Leimeister JM, Loos P (2014) Business models – an information systems research agenda. Bus Inf Syst Eng 6(1):45–53
Venkatesan R, Kumar V (2004) A customer lifetime value framework for customer selection and resource allocation strategy. J Mark 68(4):106–125
Wagner TM, Benlian A, Hess T (2014) Converting freemium customers from free to premium – the role of the perceived premium fit in the case of music as a service. Elec Mark 24(4):259–268

White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. ECTA 48(4):817–838

Wübben M, von Wangenheim F (2008) Instant customer base analysis: managerial heuristics often "get it right". J Mark 72(3):82–93

Zhang JQ, Dixit A, Friedmann R (2010) Customer loyalty and lifetime value: an empirical investigation of consumer packaged goods. J Mark Theory Pract 18(2):127–140