Available online at www.sciencedirect.com

ScienceDirect

Journal of Interactive Marketing 27 (2013) 185 – 208

JOURNAL OF INTERACTIVE MARKETING

www.elsevier.com/locate/intmar

A Partially Hidden Markov Model of Customer Dynamics for CLV Measurement

Jaime Romero [a,*] & Ralf van der Lans [b] & Berend Wierenga [c]

[a] *Universidad Autónoma de Madrid. Facultad de CC. Económicas y Empresariales. Avda. Francisco Tomás y Valiente, 5. Madrid 28049 Spain*
[b] *The Hong Kong University of Science and Technology. Clear Water Bay, Kowloon, Hong Kong*
[c] *Rotterdam School of Management, Erasmus University. Burgemeester Oudlaan, 50. 3062 PA Rotterdam The Netherlands*

Available online 27 June 2013

Abstract

Customer lifetime value (CLV) measurement is challenging as it requires forecasting customers' future purchases. Existing stochastic CLV models for this purpose generally make the following assumptions: 1) purchase behavior of customers can be described by purchase frequency and the average monetary value of transactions, 2) customers keep the same purchase behavior pattern over time, 3) purchase frequency and monetary value are independent, and 4) customers are active during a limited period of time after which they permanently defect. We develop a new stochastic model that relaxes these four assumptions. First, in addition to the number of transactions and its monetary values, we also model purchase incidence decisions (i.e. whether or not to purchase). Second, our partially hidden Markov truncated–NBD-GG (PHM/TNBD-GG) model allows dynamic purchase patterns, dependence between purchase frequency and monetary value, and customers to become active after a few periods of temporary inactivity. Validation of our model on two datasets demonstrates that if assumptions 1 to 4 of existing stochastic models are violated our model produces more accurate forecasts of future customer behavior.

Keywords: Customer lifetime value; Forecasting; Hidden Markov model

Introduction

Over the last ten years, there has been an increase in the interest in customer relationship management (CRM) by both academics and practitioners (Payne and Frow 2005). In a CRM framework, instead of transactions, a company's relationships with customers are the focus of marketing actions. Customer lifetime value (CLV), or the net present value of the future expected profit stream from a given customer, is the key metric in CRM (Blattberg, Kim, and Neslin 2008). CLV calculations of individual customers assist companies to manage relationships in terms of how to initiate, maintain and terminate them. For example, a retailer may decide to increase the service for customers with a high CLV, and target customers with a low CLV in a direct mailing.

The key challenge of CLV measurement is the development of a model to forecast the future flow of profits that each customer will provide to the company. Academics have developed a variety of models for this purpose that hold for specific contexts. These contexts are generally divided in contractual and non-contractual settings (Fader and Hardie 2009; Reinartz and Kumar 2000). In a contractual setting, the firm observes customer defection, while in a noncontractual setting defection is unobserved. Although both situations are challenging, the latter is more difficult because it involves customer decisions that are unobserved to the company. In this research, we focus on noncontractual settings, but, as we show in the discussion, our modeling approach can be easily adapted to contractual settings.

CLV models generally focus on the frequency and/or recency of purchases and the average monetary value in each purchase. Many approaches relate these quantities to managerially actionable variables (e.g. Rust, Lemon, and Zeithaml 2004; Venkatesan and Kumar 2004) and allow analyzing the impact of these variables on expected profitability. Other approaches model these

* Corresponding author.
E-mail addresses: jaime.romero@uam.es (J. Romero), rlans@ust.hk (R. van der Lans), bwierenga@rsm.nl (B. Wierenga).

1094-9968/$ -see front matter © 2013 Direct Marketing Educational Foundation, Inc. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.intmar.2013.04.003

quantities using assumptions of the underlying stochastic buying behavior of customers and usually omit explanatory covariates (Fader, Hardie, and Lee 2005b). The advantages of these stochastic models are that they are based on a simple "story" about customer behavior that fits almost every business, and that it generates a deeper understanding of how such behavior is related to CLV. Previous research has shown that these stochastic models are appropriate for addressing a number of managerial issues, such as customer selection for resource allocation or outsourcing, and CLV measurement for long periods of time. One of the first models to conduct this type of analysis is the Pareto/NBD-model by Schmittlein, Morrison, and Colombo (1987). Since their pioneering work, other studies have proposed variations of the original model (e.g.: Abe 2009; Batislam, Denizel, and Filiztekin 2007; Colombo and Jiang 1999; Fader, Hardie, and Lee 2005a,b; Fader, Hardie, and Shang 2010; Ho, Park, and Zhou 2006; Jerath, Fader, and Hardie 2011; Reinartz and Kumar 2003; Schmittlein and Peterson 1994).

Stochastic models have performed well in several business settings (e.g.: medical offices, attorneys and insurance related firms in Schmittlein and Peterson 1994; customer database selling in Colombo and Jiang 1999; catalog retailing and high-tech B2B in Reinartz and Kumar 2003, and CD retailing in Fader, Hardie, and Lee 2005a,b; Abe 2009). Recently however, some researchers have criticized the performance of these models, because relatively simple managerial heuristics lead to similar predictions in several business settings (Wübben and Wangenheim 2008). A likely explanation for this finding is that the underlying assumptions of these stochastic models may not hold or are too strict in some business settings. In particular, these stochastic models assume that: 1) purchase behavior of customers can be described by two statistics, i.e. purchase frequency and the average monetary value of these transactions, 2) customers keep the same purchase behavior pattern over time, 3) purchase frequency and monetary value are independent, and 4) customers are active during a limited period of time after which they permanently defect.

The goal of this paper is to develop a model that is more general and less restrictive. Therefore, based on previous stochastic models of repeat buying behavior, we present a model that in addition to the statistics (1) purchase frequency and (2) average monetary value, also uses (3) purchase incidence (i.e.,whether or not to purchase in a specific time period) to describe purchase behavior; (4) incorporates dynamic purchase patterns; (5) handles situations in which purchase frequency and monetary value are related; and (6) allows consumers to come back after temporary inactivity. To develop our model, we slightly change the behavioral "story" of repeat purchasing of consumers. More specifically, we assume that consumers switch their purchase behavior over time according to a partially-hidden Markov (PHM) model, in which consumers may alternate between activity and inactivity states until they permanently defect. Application of our model to two empirical settings shows that our model produces more accurate forecasts than benchmarks models widely used in the literature and practice, especially when there are indications that their assumptions, as described above, do not hold.

This paper is organized as follows. We first discuss existing stochastic models of repeat purchase behavior, after which we present our model and derive its expressions. Then, we present the empirical setting to which we validate our model and compare it to several benchmark models. In the concluding section we discuss our findings and provide implications of our research.

Stochastic Models of Repeat Purchase Behavior for CLV Measurement

Stochastic models for CLV measurement typically assume that customers make purchases until they permanently defect. These models focus on predicting purchase frequency in the future and supplement these predictions with models that forecast the monetary value of each transaction, independent from purchase frequency.

Modeling Purchase Frequency

For purchase frequency, two main models have been developed: the Pareto/NBD by Schmittlein, Morrison, and Colombo (1987) and the BG/NBD by Fader, Hardie, and Lee (2005a). Both models assume that customers purchase until they defect. The purchase and defection processes are independent. In both models, customer purchases follow a negative binomial distribution (NBD), conditional on being alive. The NBD model arises from assuming that purchases of a customer i are distributed according to a Poisson process with rate λ_i, which varies across the population of customers following a gamma distribution. Hence the NBD model accounts for customer heterogeneity in terms of purchase rates. The Pareto/NBD model assumes that customers may defect at any point in time between their last purchase and the end of the observation period, which follows a Pareto of the second-kind distribution. In contrast, the BG/NBD model assumes that customers may only defect immediately after their last purchase, according to a beta-geometric (BG) distribution. Both models are similar in their behavioral assumptions and produce comparable forecasts.

Modeling Monetary Value

Academic literature has provided several solutions to model monetary value of purchases. The most relevant ones are the models of Schmittlein and Peterson (1994), Colombo and Jiang (1999) and Fader, Hardie, and Lee (2005b). These studies provide estimates of the expected monetary value of future transactions independent of purchase frequency. Schmittlein and Peterson (1994) argue that the average monetary value of past transactions by a consumer is a reliable estimate of the transaction value of future purchases by that same customer. Therefore, Schmittlein and Peterson (1994) assume that 1) transaction value of each purchase of a customer i is independent and normally distributed with mean M_i, 2) M_i varies across customers following a normal distribution with mean M, and 3) M_i is assumed to be independent of the purchase frequency and defection processes. These assumptions allow estimating the average value of future transactions of each customer, which is a weighted average of the transaction values of past purchases of the customer and the average transaction value of the population. Particularly, the longer the purchase history of a consumer, the more weight these purchases receive. Colombo and Jiang (1999) and Fader, Hardie, and Lee (2005b) argue that the normal distribution is not appropriate for modeling transaction value – the normal distribution ranges between −∞ and +∞, while customer transaction value is strictly positive. They propose to use a gamma distribution instead of the normal distribution in assumptions 1 and 2, which runs on the interval (0, +∞]. This gamma–gamma (GG) model can be combined with purchase frequency models in order to calculate CLV. For instance, Fader, Hardie, and Lee (2005b) combine the Pareto/NBD and the GG models to produce CLV forecasts.

In sum, stochastic models for CLV calculation have a long history that started with the model of Schmittlein, Morrison, and Colombo (1987). Based on this early model, Fader, Hardie, and Lee (2005a) built the BG/NBD-GG that can be applied in a broad setting. However, like its predecessors, this model is based on behavioral assumptions that may lead to inaccurate predictions if these assumptions do not hold. The first limitation is the permanent defection assumption. Obviously, in markets where customers return to activity or follow 'always-a-share' purchase patterns, these models underestimate CLV (Rust, Lemon, and Zeithaml 2004). As a second limitation, customer dynamics are almost neglected in these models. The only change that can occur is from activity to permanent defection. An extensive body of research in marketing reveals that, often, purchase behavior is not stable across time. Sometimes this lack of stability can be explained by controllable factors, such as loyalty programs and other marketing actions. In other cases, these changes cannot be controlled by firms. For example, in some product categories we know that customers are variety seeking (Bawa 1990; Park and Gupta 2011) or that customers change behavior according to stages of their life cycle (Du and Kamakura 2006). In situations where dynamics in purchase patterns occur, these stochastic models lead to inaccurate CLV estimates. Only a few studies (Allenby, Leone, and Jen 1999; Ascarza and Hardie 2012; Ho, Park, and Zhou 2006; Schweidel, Bradlow, and Fader 2011) have incorporated dynamic purchase patterns in CLV models. Ho, Park, and Zhou (2006) allow the time to the next purchase to depend on a customer's satisfaction with the last purchase. This model, however, has not been empirically validated yet. Furthermore, Allenby, Leone, and Jen (1999) incorporate dynamics by allowing consumers to switch between periods of different interpurchase times. Their model, however, is situated in a contractual, always-a-share setting in which defection is assumed to be observed (see also Venkatesan and Bohling 2007). Schweidel, Bradlow, and Fader (2011) in a contractual setting, analyze customer evolution in terms of the portfolio of services they purchase from a multi-service provider. Ascarza and Hardie (2012), also in a contractual setting, propose that renewal decisions depend on customer latent commitment.

The third limitation is that monetary value and purchase frequency are considered to be independent, which is also not realistic in most cases. For example, Borle, Singh and Jain (2008) show in a situation where customer defection is observed that purchase frequency is negatively related to monetary value. They therefore develop a Hierarchical Bayesian model that allows purchase frequency, monetary value and defection to be related. However, their model assumes that defection is observed, which is often not the case. In the next section we present our partially-hidden Markov model (PHM) that overcomes these limitations of existing stochastic models discussed before.

Partially-Hidden Markov Truncated NBD-GG (PHM/TNBD-GG) Model

Model Assumptions

Following previous stochastic models on repeat purchases, we formulate the behavioral story that underlies our model. A key difference between our model and previous stochastic models, such as the Pareto/NBD and BG/NBD, is that we allow consumer purchase patterns to vary over time. We assume that these changes occur at discrete moments in time, but between these moments we model purchase behavior in continuous time. Similar to Allenby, Leone, and Jen (1999), we divide the observation period $[t_0..t_{end}]$ into n different periods: $t = 1$: $[t_0..t_1]$, $t = 2$: $(t_1..t_2]$, .., $t = n$: $(t_{n-1}..T]$, where t_0 represents the start and T the end of the observation period. Following Allenby, Leone, and Jen (1999), we assume that these time periods correspond to weeks, which avoids day of week effects and also corresponds to the timing of managerial decisions. Using these n different time periods, our model is built on the following assumptions (Table 1 summarizes the variables and parameters of our model):

1. In each time period t, a customer is active, temporarily inactive or permanently defected. When a customer is active s/he follows a continuous time purchase pattern in which s/he

Table 1
Summary of model variables and parameters.

Parameter/ variable	Description
K	Number of states that represent purchase activity patterns
x_{ikt}	Purchase frequency of customer i who follows a purchase pattern k in period t, according to a Poisson process with rate λ_{ikt}
m_{ikt}	Average monetary value of a customer i's purchases who follows a purchase pattern k in period t, according to a gamma distribution with parameters u_k and γ_{ikt}
λ_{ikt}	Purchase rate parameter that varies across consumers i, time periods t, and activity states k. This variation is captured through a modified gamma distribution with parameters r_k and α_k depending on activity state k
γ_{ikt}	Scale parameter of m_{ikt} that varies according to a gamma distribution with parameters w_k and δ_k.
φ	Vector that represents the initial state probabilities, i.e. the probability of being in state 1,2, …,K + 2 in period 1.
Φ	Transition matrix containing the Markov switching probabilities $\phi_{k,k'}$
p_i	Vector that represents the initial state probabilities of customer i.
$q_{ikk'}$	Probability of a customer i who is in state k in period t switches to state k' in period $t + 1$.

makes at least one purchase. When a customer is inactive or defected the customer doesn't make any purchase. However, an inactive customer still can become active, while a defected customer is permanently lost and will never make any new purchase.

2. While active in a time period t, customers make their purchases according to one out of K possible continuous time unobservable purchase patterns. Including the inactive and defection state, in every time period, each customer is in one out of $K + 2$ states.
3. Given activity or inactivity in period t, a customer i may switch to another state with probability $q_{i,k_t k_{t+1}}$, where k_t indicates the state in period t. However, a customer in the defection state (state $K + 2$) will stay in the defection state forever, i.e. $q_{i,K+2,K+2} = 1$.
4. While active in time period t, customers make their purchases in continuous time following a Poisson process with rate λ_{ikt}. Because we assume that customers make at least one purchase in active time periods, purchase frequency follows a zero-truncated Poisson distribution with parameter λ_{ikt}, which may vary across customers i, activity states k, and periods t.
5. The purchase rate λ_{ikt} is not the same for every customer i in period t that follows a purchase pattern k – it is distributed following a *modified* (to be explained later) gamma distribution with parameters r_k and α_k across the population of customers and time periods.
6. When active in a time period t, the average monetary value of a customer i's purchases m_{ikt} is gamma distributed with parameters u_k and γ_{ikt} depending on activity state k.
7. γ_{ikt} varies according to a gamma distribution with parameters w_k and δ_k.
8. Purchase frequency and monetary value are assumed to be independent within activity states, however, across states purchase frequency and monetary value may be related.

Assumptions 1 to 3 lead to a partially-hidden Markov (PHM) model in which consumers may switch in each time period between K activity states, a temporary inactivity state and the permanent defection state. Consequently, we assume all customers that are not defected to first decide whether they will purchase or not (i.e., become active or temporarily inactive), and given activity they decide which purchase pattern they will follow (i.e., one out of K activity states). This approach is consistent with choice models that model purchase incidence and frequency as independent processes (Bucklin, Gupta, and Siddarth 1998). Moreover, it allows distinguishing between a temporary inactivity state and a permanent defection state, which is a key managerial insight (Schweidel, Bradlow, and Fader 2011). Assumptions 1 to 3 imply that our model contains both observed as well as latent states. For instance, if a customer i makes a purchase in period t_n after some non-purchase periods, we observe that this customer was in the inactivity state during these non-purchase periods and is in a latent activity state in period t_n. If after t_n the customer doesn't make a purchase, we don't observe if the customer is either temporarily inactive or permanently defected.

Models that combine both latent and observed states in different time periods are called partially-hidden Markov (PHM) models (Scheffer, Decomain, and Wrobel 2001; Thompson, Thomson, and Zheng 2007; Truyen et al. 2008). Such models have been successfully applied in hydrology and machine learning. For instance, Thompson, Thomson, and Zheng (2007) model the dynamics of daily rainfall by specifying three rainfall states. The first state is observed and represents no precipitation, while the other two states are latent and represent different rainfall generation mechanisms (light and heavy), related to the amount of rain fallen in a particular area and period of time. In machine learning, PHM models are applied in situations where some of the latent states are explicitly coded by human observers, and thus treated as "observed" (e.g., Scheffer, Decomain, and Wrobel 2001).

While previous CLV literature in marketing does not deal with partially-hidden Markov models, both Markov and hidden Markov models have been successfully applied to CLV models. For instance, Berger and Nasr (1998) and Pfeifer and Carraway (2000) use a Markov model in which the probability to make a purchase in the next period is dependent on how many periods ago the consumer made the most recent purchase. In a different approach, Rust, Lemon, and Zeithaml (2004) introduce a Markov brand switching matrix as a component of their model describing how marketing investments increase CLV and subsequently return on marketing investments. Hidden Markov models have been applied by Netzer, Lattin, and Srinivasan (2008) to model donation of university alumni, in which the choice to donate depends on latent relationship states that follow a Markov chain. In addition, Ho, Park, and Zhou (2006) extend the Pareto/NBD by incorporating a two-state Hidden Markov model component representing satisfied and dissatisfied customers. Schweidel, Bradlow, and Fader (2011) propose a model of portfolio choice in multiservice contractual settings, which accounts for portfolio inertia and service stickiness. Ascarza and Hardie (2012) incorporate commitment states as a determinant of the renewal process in contractual settings. Finally, Ma and Büschken (2011) use a two-state hidden Markov model to allow customers to switch between activity and inactivity states. (For a more general discussion of HMM in these settings, see Schwartz, Bradlow, and Fader 2012).

Assumptions 4 and 5 assure that purchase frequency is distributed following a zero-truncated Poisson distribution with interpurchase rate parameter λ_{ikt} that varies across consumers i, time periods t, and activity states k. However, instead of using a gamma distribution to account for heterogeneity in purchase frequency λ_{ikt} (Fader, Hardie, and Lee 2005a, Schmittlein, Morrison, and Colombo 1987), we assume a modified gamma distribution with parameters r_k and α_k depending on activity state k. The modified gamma distribution is an adapted version of the gamma distribution and is defined as follows:

$$P(\lambda_{ikt}|r_k,\alpha_k) = \frac{(\alpha_k+1)^{r_k}}{\left((\alpha_k+1)^{r_k}-\alpha^{r_k}\right)}\left(1-e^{-\lambda_{ikt}}\right) \times gamma(\lambda_{ikt}|r_k,\alpha_k) \;\; \forall \lambda_{ikt}>0, \quad (1)$$

where $gamma(\lambda_{ikt}|r_k, \alpha_k)$ represents the standard gamma distribution. The modified gamma distribution (1) is very similar to the standard gamma distribution (see Fig. 1) and is chosen here because of the combination with the zero-truncated Poisson distribution. As derived in Appendix A, Assumptions 4 and 5 lead to a zero-truncated negative binomial distribution (zero-truncated NBD) for the number of purchases for a randomly chosen customer in activity state k (i.e., to a NBD distribution whose domain is restricted to strictly positive integers, excluding zero).

Assumptions 6 and 7 follow Fader, Hardie, and Lee (2005b) and lead to a gamma–gamma (GG) distribution for monetary value, given activity state k and time period t. Similarly with previous work (Fader, Hardie, and Lee 2005b; Schmittlein and Peterson 1994), given the activity state for a specific time period t, Assumption 8 assures that purchase frequency and monetary value are independent within states. However, we do not make any assumptions about purchase frequency and monetary value across different activity states. Therefore our approach allows purchase frequency and monetary value to be related across activity states, while keeping independence within activity states, which is computationally attractive. Note that in addition to allowing purchase frequency and monetary value to be related across different customers (Borle, Singh and Jain 2008), our formulation also allows these values to be correlated within the same customer across different time periods. The latter is because we allow purchase patterns of customers to be different across time periods. For instance, it allows customers to spend small amounts in periods with high frequency and large amounts in time periods with low purchase frequency.

In sum, our proposed partially-hidden Markov model (PHM/TNBD-GG), with a zero-truncated NBD for purchase frequency and a gamma–gamma for monetary value is the first stochastic CLV model that explicitly models purchase incidence, in addition to purchase frequency and monetary value. Our approach to modeling purchase incidence, in addition to dynamic purchase patterns, leads to a partially-hidden Markov model, which is new to the marketing literature and relaxes the independence between purchase frequency and monetary value. Furthermore, the incorporation of heterogeneity in purchase frequency through our proposed alternative gamma distribution is, to the best of our knowledge, new and accommodates probability models that try to incorporate a zero-truncation in the stochastic arrival process. The next section explicitly derives the likelihood specification of our model and shows how to estimate it.

Model Specification and Estimation

To be able to estimate the model, we first need to specify the model likelihood. Based on the assumptions specified in the previous section, the likelihood of observing a specific purchase pattern (x_i,m_i), with x_i and m_i representing ($T \times 1$) vectors of, respectively, purchase frequency and average monetary value of an individual customer i, is specified as follows:

$$L(\lambda_i, \gamma_i, u, \varphi, \Phi | x_i, m_i) = \sum_{k_1=1}^{K+2}\sum_{k_2=1}^{K+2}\cdots\sum_{k_T=1}^{K+2}\left[p_{ik_1}\cdot f_{k_1}(\lambda_{ik_1 1}|x_{i1})\cdot g_{k_1}\left(\gamma_{ik_1 t}, u|m_{i1}\right) \cdot \prod_{t=2}^{T}\left(f_{k_t}(\lambda_{ik_t t}|x_{it})\cdot g_{k_t}(\gamma_{ik_t t}, u|m_{it})\cdot q_{ik_{t-1}k_T}\right)\right]. \quad (2)$$

This likelihood function captures all the possible paths across states that a customer i might have followed from periods 1 to T, given customer i's purchase pattern. In Eq. (2), $f_k(\cdot)$ represents the zero-truncated Poisson distribution, and $g_k(\cdot)$ the gamma distribution for k = 1,.., K activity states (see Assumptions 4 and 6), and $f_{K_t+1}(\cdot) = f_{K_t+2}(\cdot) = g_{K_t+1}(\cdot) = g_{K_t+2}(\cdot) = 1$ if x_{it} = 0, zero otherwise. State K + 1 corresponds to temporary inactivity, while state K + 2 corresponds to permanent defection. The (K + 2) × (K + 2) matrix Q_i represents the Markov switching probabilities $q_{ikk'}$, and the (K + 2) -vector P_i represents the initial state probabilities. Assumption 3 implies that $q_{i,K+2,K+2}$ =1 and $q_{i,K+2,k} = 0 \;\forall\; k = 1,2,\ldots, K+1$, i.e. customers cannot leave the permanent defection state.

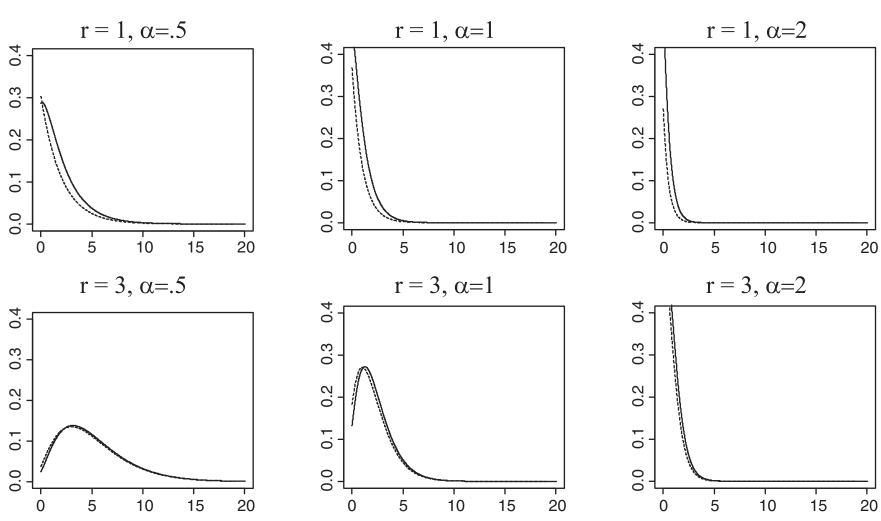

Fig. 1. Comparison of the gamma and the modified gamma distributions.

To estimate the model, we follow Fader, Hardie, and Lee (2005a) and derive the likelihood for a random consumer using likelihood (2) and Assumptions 5 and 7. In order to do so, we take expectations of the individual level parameters λ_i and γ_i for each time period and each state to arrive at closed-form model likelihood. As derived in the previous paragraph, this leads to zero-truncated NBD distributions for purchase frequency, and gamma–gamma distributions for monetary value in each time period. Hence, the likelihood for a randomly chosen customer with $T \times 1$ -vectors of purchase frequency x and monetary value m is defined as follows:

$$L(r,\alpha,w,\delta,u,\varphi,\Phi|x,m) = \sum_{k_1=1}^{K+2}\sum_{k_2=1}^{K+2}\cdots\sum_{k_T=1}^{K+2}\Big[\varphi_{k_1}\cdot h_{k_1}(r_{k_1},\alpha_{k_1}|x_1)\cdot \ell_{k_1}(w_{k_1},\delta_{k_1},u_{k_1}|m_1) \cdot\prod_{t=2}^{T}(\phi_{k_{t-1}k_t}\cdot h_{k_t}(r_{k_t},\alpha_{k_t}|x_t)\cdot \ell_{k_t}(w_{k_t},\delta_{k_t},u_{k_t}|m_t))\Big], \quad (3)$$

with $h_{k_t}(\cdot)$ representing the zero-truncated NBD, and $\ell_{k_t}(\cdot)$ the gamma–gamma distribution for $k_t = 1,.., K$, and $h_{K_t+1}(\cdot) = h_{K_t+2}(\cdot) = \ell_{K_t+1}(\cdot) = \ell_{K_t+2}(\cdot) = 1$ if $x_t = 0$, zero otherwise. Parameters φ_k and $\phi_{k,k'}$ represent, respectively, the overall initial state probabilities and Markov switching probabilities.

To estimate the model parameters $\Theta = \{\varphi,\Phi,r,\alpha,u,w,\delta\}$ in likelihood (3), we built upon the Expectation Maximization (EM) algorithm for hidden Markov models (Rabiner 1989) that results in efficient estimators (Ephraim and Merhav 2002), and included two modifications. First, due to our treatment of inactivity as zero purchases that are observed, we followed Zhang and Mason (1989) and introduced a supervision function that modifies the probabilities of being in one of the $K + 2$ states. For every customer i and time periods $t = 1, \ldots, T$, this supervision function examines the purchase pattern and assigns 'observed' probabilities to state memberships. If customer i makes more than 1 purchase in period t, the probability of being inactive or defected in that period is set to zero. Otherwise, if customer i does not purchase in period t, the probability of being in one of the K activity states is set to 0. Second, because the EM algorithm for hidden Markov model estimation was originally developed for speech recognition systems in which data input is a long series of observations from one subject, we followed Li, Parizeau, and Plamondon (2000) to apply the algorithm to relatively short series of many individuals. In particular, we apply the EM algorithm at an individual level, for every subject, and aggregate the results by averaging the individual parameter estimates at each iteration (see Appendix B for the details). This results in the estimation of the overall initial probabilities φ_k and Markov switching probabilities $\phi_{k,k'}$, in which the corresponding individual-specific parameters (i.e., p_{ik} and $q_{ikk'}$) are generated as well. To determine the number of latent activity states K, we estimated the model for different values of K and selected the best fitting model using the BIC criterion. To avoid local optima, we estimate the model 200 times with different random starting values and select the solution with maximum likelihood value. Application of our model to synthetic data showed that our model was able to recover all parameters within a confidence interval of 95%.

Forecasting Purchase Frequency and Monetary Value

To calculate the CLV of a specific customer i, we need to be able to predict purchase frequency and monetary value of this customer. Following Fader, Hardie, and Lee (2005a), we use Bayes' theorem to arrive at a weighted average between a customer's previous observations and the overall population mean to infer the expected purchase frequency and monetary value. Using Bayes' rule, given a randomly chosen customer making $x_{it} > 0$ purchases in activity state k, the expected purchase frequency for this customer in state k in a time period t is (see Appendix C for the derivation):

$$E(X_{ikt}|r_k,\alpha_k,x_{it}) = (\alpha_k+1)^{r_k+x_{it}}\cdot(x_{it}+r_k)\cdot\zeta(x_{it}+r_k+1,\alpha_k+1), \quad (4)$$

where $\zeta(.)$ is Hurwitz zeta function (Hurwitz 1882; see Appendix C for the details). Like the Gaussian hypergeometric function, which is central to forecasting purchase frequency in the BG/NBD model (Fader, Hardie, and Lee 2005a), evaluation of Hurwitz zeta function is straightforward and can be approximated with a polynomial series. Similarly, Fader, Hardie, and Lee (2005b) present the expected monetary value M_{ikt} for a randomly chosen customer:

$$E(M_{ikt}|u_k,w_k,\delta_k,m_{it},x_{it}) = \frac{w_k-1}{u_k x_{it}+w_k-1}\cdot\frac{\delta_k u_k}{w_k-1} + \frac{u_k x_{it}}{u_k x_{it}+w_k-1}m_{it}. \quad (5)$$

Because we do not observe for each customer i in which state k purchases in time period t are generated, we compute their probabilities p_{ikt}, which come naturally available in the EM algorithm (see B5 in Appendix B). Therefore to forecast the expected purchase frequency X_{ik} and monetary value M_{ik} for customer i in state k, we compute a weighted average of Eqs. (4) and (5). Hence:

$$E(X_{ik}|r_k,\alpha_k,x_i,p_{ik}) = \sum_{t=1}^{T}p_{ikt}\cdot E(X_{ik}|r_k,\alpha_k,x_i)\Big/\sum_{t=1}^{T}p_{ikt}, \quad \text{and} \quad (6)$$

$$E(M_{ik}|u_k,w_k,\delta_k,m_i,x_i,p_{ik}) = \sum_{t=1}^{T}p_{ikt}E(M_{ikt}|u_k,w_k,\delta_k,m_{it},x_{it})\Big/\sum_{t=1}^{T}p_{ikt}. \quad (7)$$

In order to forecast purchase frequency and monetary value, we need in addition to Eqs. (6) and (7) to be able to predict the probabilities that customer i is in state k in period $T + \tau$, with $\tau > 0$. As provided in Appendix B (Equation B5), the EM-algorithm automatically produces $E(p_{ikT})$, i.e. the expected probability of customer i being in state k in period T, and $E(q_{ijk})$, i.e. the expected probability that customer i switches from state j to k (see

B11 in Appendix B). Hence, the predicted probabilities that customer i being in each state k in period $T+\tau$ equal:

$$E(p_{i,T+\tau}) = \left(\prod_{t=T+1}^{T+\tau} q_i\right)\cdot p_{iT}, \tag{8}$$

where p_{iT} is a $(K+2)$ -vector containing the probabilities $E(p_{ikT})$, and $q_i = E(Q_i)$ a $(K+2)\times(K+2)$ matrix containing the switching probabilities $E(q_{ijk})$. Note that using Eq. (8), it is straightforward to compute the probability that a specific customer makes a purchase in a specific future time period. This quantity is of great managerial importance (Netzer, Lattin, and Srinivasan 2008), but is difficult to compute in the Pareto/NBD and BG/NBD models (Wübben and Wangenheim 2008). Using Eqs. (6) to (8), the predicted purchase frequency and monetary value of customer i in period $T+\tau$ equal:

$$E(X_{i,T+\tau}) = E(p_{i,T+\tau})'\cdot E(X_i) \quad \text{and} \tag{9}$$

$$E(M_{i,T+\tau}) = E(p_{i,T+\tau})'\cdot E(M_i), \tag{10}$$

with $E(X_i)$ and $E(M_i)$ $(K+2)$ -vectors containing respectively expected purchase frequencies and monetary values for customer i being in each state. Similarly, given discount rate d and by defining the vector $E(V_i)$ as the vector of products of purchase frequency and monetary value in each state, the CLV of customer i can be computed as follows:

$$\text{CLV}_i = \sum_{t=T+1}^{T+\tau} E(p_{it})'\cdot E(V_i)\cdot \frac{1}{(1+d)^{t-T}}. \tag{11}$$

Empirical Applications

We validated our PHM/TNBD-GG model on two data sets (see Table 2 for descriptive statistics). The first data set contains scanner data from a hypermarket in the northeast of Spain. The second dataset consists of online CD sales and was used by Fader and Hardie (2001) and Fader, Hardie, and Lee (2005a,b), among others. These datasets are described next.

Hypermarket Dataset

This dataset contains purchase records from 3017 customers of a hypermarket in the northeast of Spain for a period of one year (January 1999 to December 1999). All the customers in our dataset have purchased at least once during the four first weeks of the year. Table 2 presents the descriptive statistics of the data for each quarter (13 weeks). Overall, given activity, customers do 1.36 shopping trips on average each week in which they spend €39.23. These numbers are relatively stable in each quarter, although the average expenditure per shopping trip is slightly increasing over time (from €37.81 in the first quarter to €41.14 in the final quarter). More interestingly, the number of customers that make at least one purchase in each week decreases over time. On average, during the first quarter, 1192 customers do at least one shopping trip every week, while in the fourth quarter this is only 905.6, which suggests that the inclusion of an absorbing defection state is important. Although many customers become inactive over time, there is a substantial group of customers that start purchasing after a week or longer of inactivity (on average 496.8 customers). This suggests that customer dynamics might be an important feature of a model that tries to capture purchase frequency and monetary value.

CDNOW Dataset

This dataset contains purchase records from 2349 customers of the online retailer CDNOW in the period from January 1997 through June 1998 (79 weeks). These data were originally used by Fader and Hardie (2001) and have later been used in several studies to validate the BG/NBD model (Fader, Hardie, and Lee 2005a; Wübben and Wangenheim 2008). We have selected

Table 2
Descriptive statistics.

Weeks	Average nr. of active customers by week	Average nr. of inactive customers by week	Average nr. of customers becoming inactive or defected by week	Average nr. of customers becoming active by week	Average shopping trips by week	Average expenditure by shopping trip [1]
Hypermarket						
1–13	1192.1	1824.9	611.0	587.9	1.35	37.81
14–26	983.2	2033.8	489.9	482.9	1.36	39.03
27–39	940.3	2076.7	476.5	470.4	1.37	39.09
39–51	905.6	2111.4	459.9	449.2	1.36	41.14
1–51	1007.2	2009.8	508.3	496.8	1.36	39.23
CDNOW						
1–13	231.1	2117.9	198.7	212.4	1.08	33.11
14–26	67.2	2281.8	59.6	54.2	1.07	36.56
27–39	51.3	2297.7	44.4	41.5	1.12	36.07
39–52	53.3	2295.5	42.4	43.8	1.11	36.99
53–65	48.2	2300.85	36.2	38.1	1.09	36.95
66–79	34.8	2314.2	32.2	30.3	1.09	35.92
1–79	80.4	2268.6	73.1	71.8	1.09	34.91

[1] Average monetary value in the hypermarket dataset is measured in Euros, while for the CDNOW dataset this is measured in US Dollars.

Table 3
Parameter estimates for the hypermarket data.

Parameters	Active state 1	Active state 2	Active state 3	Active state 4	Active state 5	Inactive state	Defection state
Purchase frequency							
r	3.89	3.89	3.89	3.89	3.89	–	–
α	10.37	5.28	6.81	5.60	17.95	–	–
Monetary value							
u	5.85	431.75	28.42	2592.91	377.26	–	–
w	925.65	32.95	2103.17	30.75	10.04	–	–
δ	1175.81	1.05	1705.93	0.44	1.82	–	–
Mean purchase frequency	1.25	1.50	1.38	1.47	1.14	0	0
Mean monetary value	7.44	14.17	23.06	38.31	76.00	0	0
Total expenditure	9.28	21.27	31.89	56.34	86.63	0	0
Markov switching probabilities (*from/to*)							
From active state 1	.21	.13	.08	.05	.02	.46	.04
From active state 2	.07	.23	.12	.07	.03	.44	.03
From active state 3	.03	.09	.21	.13	.04	.46	.04
From active state 4	.02	.04	.12	.26	.09	.44	.02
From active state 5	.01	.02	.04	.11	.27	.53	.03
From inactivity	.03	.05	.09	.09	.11	.62	.00
From defection	–	–	–	–	–	–	1.00
Initial probabilities	.04	.07	.11	.12	.11	.54	–
Average percentage of customers	.03	.06	.09	.10	.09	.48	.16
Percentage of customers in period T	.02	.05	.08	.09	.09	.37	.30

Note: the percentages of customers in each state are calculated as the sum of the probabilities of being in each state across customers.

customers who make at least one purchase whose monetary value is higher than zero during the first 40 weeks. Table 2 provides descriptive statistics for these data for each quarter. Overall, purchase frequency is lower compared to the hypermarket dataset (on average 1.09 shopping trips per week), while average monetary value at a shopping trip is $34.91. While monetary value slightly increased over time for customers in the hypermarket, purchase frequency and monetary value are stable in each quarter for the CDNOW data. The number of active customers in each week decreases rapidly after the first quarter

Table 4
Parameter estimates for the CDNOW data.

Parameters	Active state 1	Active state 2	Active state 3	Active state 4	Active state 5	Active state 6	Active state 7	Inactive state	Defection state
Purchase frequency									
r	2.78	2.78	2.78	2.78	2.78	2.78	2.78	–	–
α	123.81	47.27	82.51	66.43	30.22	42.93	1.47	–	–
Monetary value	2.78	2.78	2.78	2.78	2.78	2.78	2.78		
u	15.86	117.62	560.03	24.83	1181.97	985.62	7.5	–	–
w	676.72	1595.06	1337.23	1011.67	32.27	7.82	4.22	–	–
δ	288.24	158.58	35.40	1065.42	1.24	0.69	12.16	–	–
Mean purchase frequency	1.02	1.04	1.02	1.03	1.06	1.04	2.48	0.00	0.00
Mean monetary value	6.76	11.70	14.84	26.17	46.77	100.05	28.3	0.00	0.00
Total expenditure	6.87	12.17	15.18	26.92	49.72	104.49	70.14	0.00	0.00
Markov switching probabilities (*from/to*)									
From active state 1	.03	.00	.02	.02	.00	.00	.00	.70	.23
From active state 2	.00	.02	.01	.02	.01	.00	.00	.76	.17
From active state 3	.00	.01	.03	.02	.00	.00	.01	.73	.20
From active state 4	.00	.01	.01	.03	.02	.00	.01	.79	.13
From active state 5	.00	.01	.02	.02	.03	.02	.01	.76	.14
From active state 6	.00	.01	.02	.02	.02	.04	.00	.79	.11
From active state 7	.00	.03	.01	.05	.03	.02	.17	.58	.11
From inactivity	.00	.01	.01	.02	.02	.01	.00	.89	.04
From defection	–	–	–	–	–	–	–	–	1.00
Initial probabilities (t = 1)	.03	.16	.19	.34	.18	.10	.01	.00	–
Average percentage of customers	.00	.02	.02	.04	.02	.01	.00	.38	.51
Percentage of customers in period T	.00	.00	.00	.01	.01	.00	.00	.17	.80

Note: the percentages of customers in each state are calculated as the sum of the probabilities of being in each state across customers.

Table 5
Out-of-sample forecasts for hypermarket data.

Criterion	PHM/TNBD-GG	BG/NBD-GG	Pareto/NBD-GG	NBD-GG	Managerial heuristics
Individual customers measures (*MAD*)					
Total expenditure	17.44 (.090)	17.90 (.087)	17.56 (.088)	19.29 (.083)	18.66 (.086)
Purchase frequency	.44 (.0017)	.45 (.0017)	.44 (.0017)	.49 (.0016)	.47 (.0017)
Purchase incidence	.32 (.0009)	.34 (.0008)	.34 (.0009)	.38 (.0007)	.34 (.0009)
Aggregate measures (MAPE)					
Total expenditure	14.12 (1.67)	8.45 (1.44)	12.87 (1.88)	14.42 (8.37)	12.81 (1.53)
Purchase frequency	10.32 (1.99)	6.32 (.65)	10.05 (1.28)	16.48 (1.59)	15.36 (1.55)

Note: MAD = mean absolute deviation, MAPE = mean absolute percentage error, standard errors between brackets.

(231.1 active customers in the first quarter, vs. 67.2 in the second quarter, vs. 34.8 in the final quarter). Similar to the hypermarket dataset, we also observe many customers who become active after a period of being inactive (on average 71.8 customers a week), although customer defection in the CDNOW dataset is more prevalent.

Estimation Results and Model Validation

For model estimation and validation we split both data sets into two halves. The first half (hypermarket: weeks 1–24, and CDNOW: weeks 1–40) is used to estimate the model parameters, while the remaining weeks are used for model validation. For the hypermarket, all customers are existing clients of the hypermarket, and hence we start the estimation at week 1 for each customer. In contrast, for the CDNOW data we followed Fader, Hardie, and Lee (2005a,b), and started the estimation for each customer at the moment of their first purchase. Because variation in purchase frequency is relatively low within time periods, we restricted shape parameter r of the modified gamma distribution to be equal across states to empirically identify our model.

Model Selection

On both datasets, we estimated the PHM/TNBD-GG allowing for a different number of activity states. We selected the model that fitted best based on the BIC-criterion. Using this procedure, we selected a model with seven states (including inactivity and defection states) for the hypermarket data (BIC: 3 states: 344,067; 4 states: 326,702; 5 states: 316,129, 6 states: 308,522; 7 states 303,307; 8 states 304,982), and nine states for the CDNOW data (BIC: 3 states: 55,738; 4 states: 53,570; 5 states: 52,517; 6 states: 52,398; 7states: 51,750; 8 states: 50,536; 9 states: 50,022; 10 states: 51,099). Tables 3 and 4 present the parameter estimates for both datasets that we discuss next.

Estimation Results Hypermarket

Table 3 presents the parameter estimates of the five activity states for customers of the hypermarket. States 1 and 5 contain customers with a relatively low purchase frequency (respectively 1.25 and 1.14 shopping trips per week, compared to 1.50, 1.38 and 1.47 in States 2 to 4, respectively), but these two states are strongly differentiated in terms of average amount of monetary value spent in each shopping trip (respectively €7.44 and €76.00, compared to €14.17, €23.06 and €38.31 in States 2 to 4, respectively). Although not very strong, our results suggest some negative relationship between purchase frequency and monetary value, with State 5 representing customers with the lowest purchase frequency and highest mean monetary value.

The Markov switching probabilities between the five active states, inactivity and defection provide several interesting insights. Customers in all states have a moderate tendency to return to the hypermarket, as the probability of purchasing again during the next period ranges from .45 (State 5) to .53 (State 4). Thus, while customers in State 5 have the highest total expenditure, these customers tend to wait slightly longer to make a new shopping trip. The results also show that every week a substantial percentage (37 percent) of customers who are inactive return after some weeks of inactivity. Interestingly,

Table 6
Out-of-sample forecasts for CDNOW data.

Criterion	PHM/TNBD-GG	BG/NBD-GG	Pareto/NBD-GG	NBD-GG	Managerial heuristics
Individual customers measures (*MAD*)					
Total expenditure	1.18 (.024)	1.24 (.023)	1.36 (.023)	2.68 (.024)	1.51 (.024)
Purchase frequency	.03 (.0005)	.03 (.0005)	.04 (.0005)	.08 (.0005)	.04 (.0005)
Purchase incidence	.03 (.0004)	.04 (.0004)	.04 (.0004)	.15 (.0003)	.04 (.0004)
Aggregate measures (*MAPE*)					
Total expenditure	29.68 (2.50)	26.79 (2.52)	26.18 (3.70)	209.46 (18.34)	43.87 (7.00)
Purchase Frequency	33.97 (3.82)	25.86 (4.14)	25.23 (4.04)	240.56 (30.98)	42.88 (12.19)

Note: MAD = mean absolute deviation, MAPE = mean absolute percentage error, standard errors between brackets.

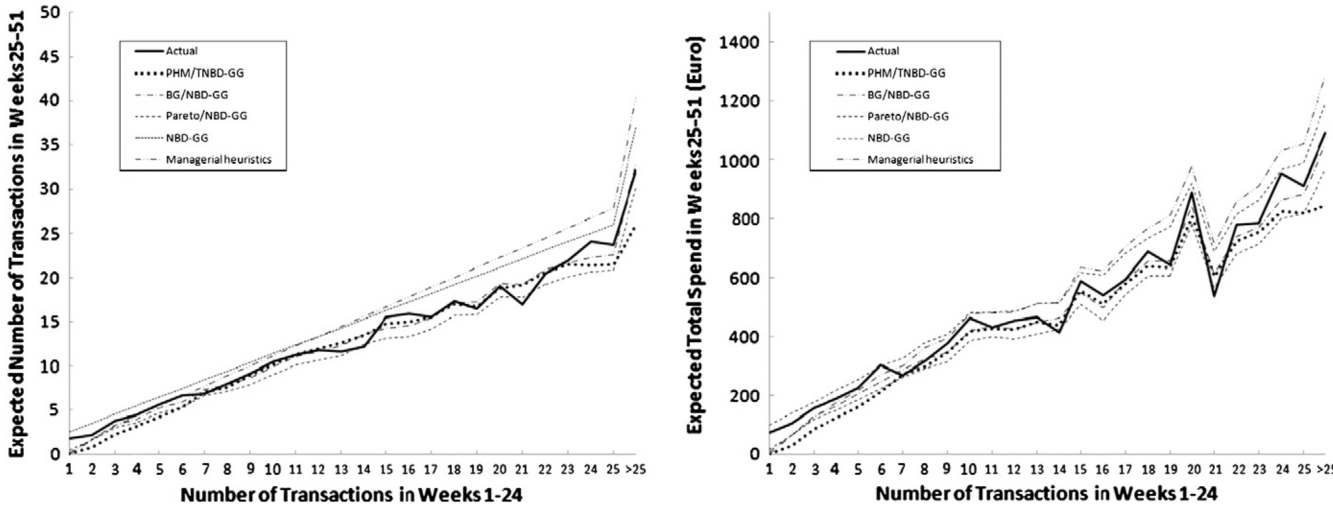

Fig. 2. Conditional expectations of purchasing for hypermarket data.

we find that defection mainly takes place after purchase activity, which suggests that the additional assumption of the BG/NBD model that consumers only defect immediately after a purchase is a reasonable one in this dataset. Furthermore, the probability of moving from an activity state to defection is quite similar across states. Interestingly, transitions between activity states evidence that customers do not drastically change their expenditure patterns over time. If customers remain active, they tend to stay in the same state or switch to a state with similar total expenditure (for instance, given switching to another activity state, customers in State 1 tend to switch to State 2). Overall, these results denote that purchase behavior in this hypermarket is relatively stable – customers alternate between activity and inactivity periods at a similar pace until they permanently defect.

Estimation Results CDNOW

Table 4 presents the results for the customers of CDNOW. States 1 to 6 have a relative low purchase frequency (between 1.02 and 1.06) but show a high variability in term of average expenditure (between \$6.76 and \$100.05). Customers in State 7 have the highest purchase frequency (2.48) with a moderate average expenditure (\$28.30). In contrast to the results of the hypermarket data, our results do not indicate a relationship between purchase frequency and monetary value. The Markov switching probabilities, however, suggest a stronger evolution in purchase patterns compared to customers of the hypermarket. While customers in the hypermarket, given remaining active, mostly stayed in the same activity state, customers of CDNOW have a relative high tendency to switch to another activity state. Exceptions are customers in state 7, which are presumably more loyal customers. However, overall the propensity to become inactive or defect is high, suggesting that CDNOW is losing customers. This is also reflected by the high percentage (80 percent) of customers in the defection state at the end of the estimation period. Interestingly, while we did not find switching from inactivity to defection for customers of the hypermarket, 4 percent of customers of CDNOW in the inactivity state switch

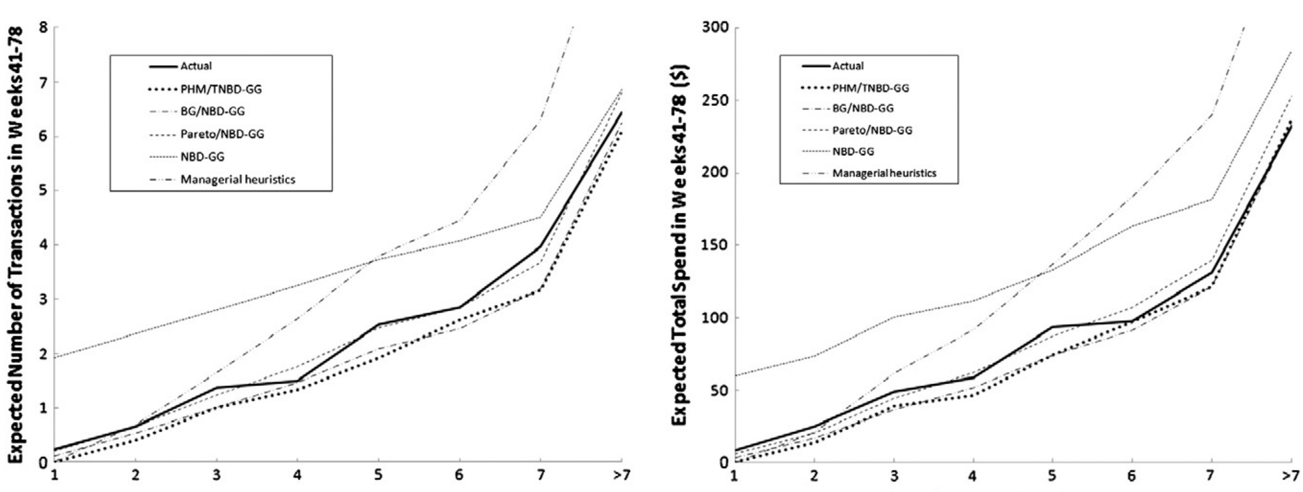

Fig. 3. Conditional expectations of purchasing for CDNOW data.

Table 7
Regression analyses of model performance comparisons with PHM/TNBD-GG.

Variable	BG/NBD-GG	Pareto/NBD-GG	NBD-GG	Managerial heuristics
Hypermarket				
# Purchases	.05***	.05***	.08***	.00
Average monetary	−.01***	−.01***	−.02***	−.03***
Inertia	.53**	.05	−.48	.36
R^2(frequency, monetary value)	−.72***	−.36**	−1.12***	−2.05***
CDNOW				
# Purchases	.04***	.03***	−.04***	−.15***
Average monetary	−.004***	−.01***	−.34***	−.01***
Inertia	.56**	.68***	−1.73***	.93*
R^2(frequency, monetary value)	−.16*	−.35***	2.95***	−1.27***

Dependent variables: MAD PHM/TNBD-GG minus MAD benchmark for each customer.
* Significant at 10% level, ** significant at 5% level, *** significant at 1% level.

to defection in every time period, which violates the BG/NBD assumption.

Model Validation

We have compared the performance of the PHM/TNBD-GG to four benchmark models. The first two are the BG/NBD and Pareto model for purchase frequency in combination with the gamma–gamma model for monetary value (i.e., BG/NBD-GG and Pareto/NBD-GG model). The third one is the standard NBD model for purchase frequency and a GG model for monetary value (NBD-GG) as discussed in Fader, Hardie, and Lee (2005b) and adapted from Colombo and Jiang (1999). Our final benchmark is a set of managerial heuristics that are frequently used in practice. More precisely, for purchase frequency and monetary value, our heuristic included the assumption that every customer will continue buying at her past mean purchase frequency and monetary value, and for defection we applied the hiatus heuristic, i.e., customers who have not purchased in a time span of length c_{hiatus} will not purchase in the future (Wübben and Wangenheim 2008). These benchmark models possess several features that make these models interesting alternatives. The managerial heuristics are popular in practice and Wübben and Wangenheim (2008) show that this approach performs often at least as well as stochastic models, making it a tough benchmark. The NBD-GG assumes that customers are always active, but does not take into account customer dynamics and assumes independence between purchase frequency and monetary value. The BG/NBD-GG and Pareto/NBD-GG models are similar to the NBD-GG model, but assume that customers permanently defect after a specific time period of activity. As explained previously, similar to the other benchmarks, these models also do not take into account customer dynamics and assume independence between purchase frequency and monetary value.

We have estimated these four benchmark models and produced predictions of the number of purchases made by each customer, their total expenditure, and purchase incidence (see Appendices D to F for the computations of purchase incidence for the first three benchmark models). For the hiatus heuristic, we followed Wübben and Wangenheim (2008) who set c_{hiatus} to six months for the CDNOW dataset. For the hypermarket dataset we did not have access to managerial information that could be used to determine c_{hiatus}. Thus, we compared several settings of the hiatus and selected the one that produced the most accurate forecasts (in this case 5 weeks), resulting in a conservative benchmark. Tables 5 and 6 present the out-of-sample forecasting results. We followed Abe (2009), and computed model fit for both individual customers (i.e., individual weekly forecasts) and for the customer base as a whole (i.e., cumulative weekly forecasts and total forecasts). For individual customers' forecasts, we computed mean absolute deviation (MAD) between predicted and observed weekly individual consumer purchase frequency, total expenditure, and purchase incidence. For the aggregate performance level, we computed mean absolute percentage errors (MAPE) between weekly observed and predicted purchase frequency and total expenditure across all consumers.

At the individual customer level, which is managerially most relevant (Wübben and Wangenheim 2008), the PHM/TNBD-GG outperforms all benchmarks[1]. For purchase frequency, performance differences between our PHM/TNBD-GG and the benchmarks, except NBD-GG, are relatively small. For total expenditure, however, these differences become larger. Apparently, taking into account a possible relationship between average monetary value and purchase frequency improves forecasting accuracy. Another important feature of our model is that it automatically computes the probability of purchase incidence in every time period, while for the first three benchmarks this requires a separate computation (see Appendices D–F). This is an important measure in practice and in some situations even more important than the total amount spent (Netzer, Lattin, and Srinivasan 2008). Tables 5 and 6 show that also for the probability of purchase incidence during a specific week, our model outperforms all four benchmarks. This suggests that a separate model for purchase incidence, next to purchase frequency and monetary value, is an important feature in predicting future purchases and incidence. Conditional expectation plots (Figs. 2

[1] Paired-samples t-tests revealed that all MADs of our proposed model are significantly lower compared to the MADs of all benchmark models ($p < .01$).

Table 8
Out-of-sample forecasts for dynamic customers of hypermarket.

Criterion	PHM/TNBD-GG	BG/NBD-GG	Pareto/NBD-GG	NBD-GG	Managerial heuristics
Individual customers measures (*MAD*)					
Total expenditure	28.01 (.31)	28.36 (.29)	28.38 (.29)	28.58 (.29)	28.20 (.30)
Purchase frequency	.61 (.0051)	.62 (.0047)	.62 (.0047)	.62 (.0047)	.61 (.0049)
Purchase incidence	.45 (.0020)	.47 (.0012)	.47 (.0012)	.48 (.0018)	.45 (.0021)
Aggregate measures (*MAPE*)					
Total expenditure	12.04 (1.87)	14.19 (2.16)	14.35 (2.17)	16.60 (2.40)	16.01 (2.34)
Purchase frequency	8.41 (1.72)	10.20 (1.97)	10.37 (1.98)	12.74 (2.17)	12.74 (2.17)

Note: MAD = mean absolute deviation, MAPE = mean absolute percentage error, standard errors between brackets.

and 3), further validate model performance at the individual customer level. Especially for the CDNOW data, both managerial heuristics as well as the NBD-GG lead to inaccurate forecasts. On the other hand, our model, as well as the BG/NBD-GG and Pareto/NBD-GG models fit the data very well.

At the aggregate level, in which we sum all weekly individual level forecasts across consumers, the BG/NBD-GG and Pareto/NBD-GG model systematically perform best (see Tables 5 and 6). The PHM/TNBD-GG seems to perform similar to managerial heuristics in both datasets, while the NBD-GG does particularly badly for the CDNOW dataset as it does not incorporate permanent defection. As in previous studies (Abe 2009; Wübben and Wangenheim 2008), our results show that models that perform best at the individual customer level do not necessarily produce better forecasts at the aggregate level.

To determine how individual differences affect model performance, we regressed the differences between MAD of total expenditure computed by the PHM/TNBD-GG and each of the four benchmarks on 1) the number of purchases in the calibration period, 2) average expenditure per shopping trip in the calibration period, 3) inertia of purchase patterns during activity, as measured by the sum of the diagonal elements of the switching matrix (i.e., $\sum_{k=1}^{K} E(q_{ikk})$), and 4) the strength of the relationship between purchase frequency and monetary value, as measured by the squared correlation between purchase frequency and monetary value in the calibration period (we set this to zero in case the correlation does not exist, or a customer made only one purchase in the calibration period). Table 7 presents the results of these regression analyses. Except for differences compared to the NBD-GG model for CDNOW, most likely due to the very bad performance of this model, the regression coefficients are very consistent. For both datasets we find negative effects for the strength of the relationship between purchase frequency and monetary value, and positive effects for inertia of purchase patterns. These results indicate that the relative performance of our model improves for consumers with dynamic purchase patterns (relative MAD increases with inertia) and when purchase frequency and monetary value are related. The effects of average expenditure are negative, while the effects of purchase frequency are mostly positive, suggesting that our model does relatively well for consumers with infrequent purchase patterns with higher monetary value.

These regression results show that our model does better for customers with dynamic purchase patterns. A possible explanation for the relatively poor performance of the PHM/TNBD-GG at the aggregate level is that purchase patterns in both datasets are relatively stable for most customers. This is not surprising for our type of empirical applications, especially for loyal customers of a hypermarket. In the datasets that we used here there is a dominance of static customers, which "hides" the advantage of using a model that allows for dynamics. The distinguishing features of our model will become more clear, also at the aggregate level if we look specifically at more dynamic and stable customers, as we discuss next.

Model Performance for Customers with Dynamic vs. Stable Purchase Patterns

To obtain a deeper comprehension of the performance of our model in situations where purchase patterns are dynamic or stable, we have re-estimated our model using four sub-samples of 250 customers[2] with the most dynamic and stable purchase patterns in both datasets. In order to select customers with dynamic (stable) purchase patterns, we have calculated the autocorrelation in total expenditure across periods for each customer and chosen those customers with the most negative (positive) autocorrelations, representing dynamic (stable) purchase patterns.

Tables 8 and 9 present the out-of-sample forecasting results of the benchmark models and our PHM/TNBD-GG for dynamic purchase patterns. For both datasets, similar to our previous results, our model outperforms benchmarks at the individual customer level. All MADs are lower than the ones from the other models, although for the hypermarket data the difference between our model and managerial heuristics is not significant[3]. Conditional expectation plots (Figs. 4 and 5) confirm these results, especially for the CDNOW dataset. More importantly, at the aggregate level we observe now similar results. While

[2] For CDNOW data, we could only select 181 customers with stable purchase patterns, because most purchase patterns consist of only few purchases.

[3] Paired-samples t-tests revealed that all MADs of our proposed model are significantly lower compared to the MADs of all benchmark models ($p < .01$), except managerial heuristics ($p > .05$) for the hypermarket dataset.

Table 9
Out-of-sample forecasts for dynamic customers of CDNOW.

Criterion	PHM/TNBD-GG	BG/NBD-GG	Pareto/NBD-GG	NBD-GG	Managerial heuristics
Individual customers measures (MAD)					
Total expenditure	4.38 (.11)	4.52 (.11)	5.70 (.11)	6.64 (.11)	5.13 (.11)
Purchase frequency	.12 (.0025)	.13 (.0025)	.16 (.0023)	.19 (.0022)	.14 (.0024)
Purchase incidence	.11 (.0020)	.12 (.0019)	.15 (.0017)	.32 (.0010)	.13 (.0019)
Aggregate measures (*MAPE*)					
Total expenditure	55.86 (7.91)	54.65 (7.15)	120.56 (14.77)	175.11 (18.90)	95.55 (12.70)
Purchase frequency	43.30 (6.20)	44.54 (5.97)	114.75 (12.12)	169.19 (15.29)	86.84 (10.39)

Note: MAD = mean absolute deviation, MAPE = mean absolute percentage error, standard errors between brackets.

previously the BG/NBD-GG and Pareto/NBD-GG models demonstrated best performance, the PHM/TNBD-GG model outperforms all benchmarks, except for MAPE of total expenditure compared to the BG/NBD-GG model in the CDNOW dataset. These differences are also relatively large: for total expenditure in the hypermarket dataset MAPE improves by at least 15 percent (12.04 vs. 14.19, respectively for PHM/TNBD-GG and the best performing benchmark BG/NBD-GG). For CDNOW, although the BG/NBD-GG does slightly better for total expenditure (MAPE: 55.86 vs. 54.65), the difference with the other models are large, with managerial heuristics having the next best performance (MAPE: 95.55). These differences in performance at the aggregate level are further illustrated in the plots that track cumulative transactions (Figs. 6 and 7). Clearly, the PHM/TNBD-GG is closest to actual cumulative purchases.

Tables 10 and 11 present the out-of-sample forecasting results of the benchmark models and our PHM/TNBD-GG for stable purchase patterns. As predicted by the results of the regression, the PHM/TNBD-GG model is not the best performing model anymore at the individual level. In the hypermarket dataset, the Pareto/NBD-GG model performs best, while in the CDNOW data the BG/NBD-GG model has the best performance at the individual level. Interestingly, the performance of the PHM/TNBD-GG is ranked second in both datasets, suggesting that the performance of our model is still satisfactorily when purchase patterns are relatively stable. More surprisingly, at the aggregate level our model has the best performance, except for purchase frequency in the CDNOW dataset. This again shows that high performance at the individual level does not guarantee optimal performance at the aggregate level, a result that was also obtained by Abe (2009) and Wübben and Wangenheim (2008).

Discussion

Since the seminal work by Schmittlein, Morrison, and Colombo (1987) and follow up research by Fader, Hardie, and Lee (2005a,b), stochastic models of customer behavior have extensively proven their usefulness in CLV modeling. However, due to their sometimes stringent behavioral assumptions, these models have also been criticized by researchers showing that simple managerial heuristics produce similar forecasting results (Wübben and Wangenheim 2008). The PHM/TNBD-GG model that this research introduces alleviates some limitations of previous stochastic models. Compared to these models of customer purchase behavior, our approach explicitly models the purchase incidence decision, next to

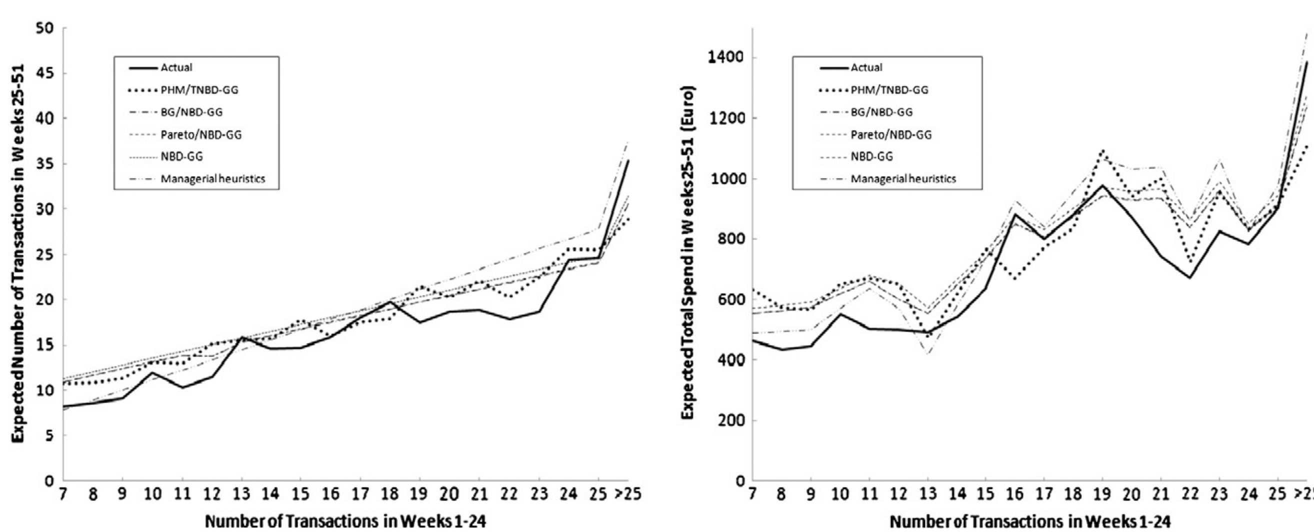

Fig. 4. Conditional expectations of purchasing for dynamic customers of hypermarket.

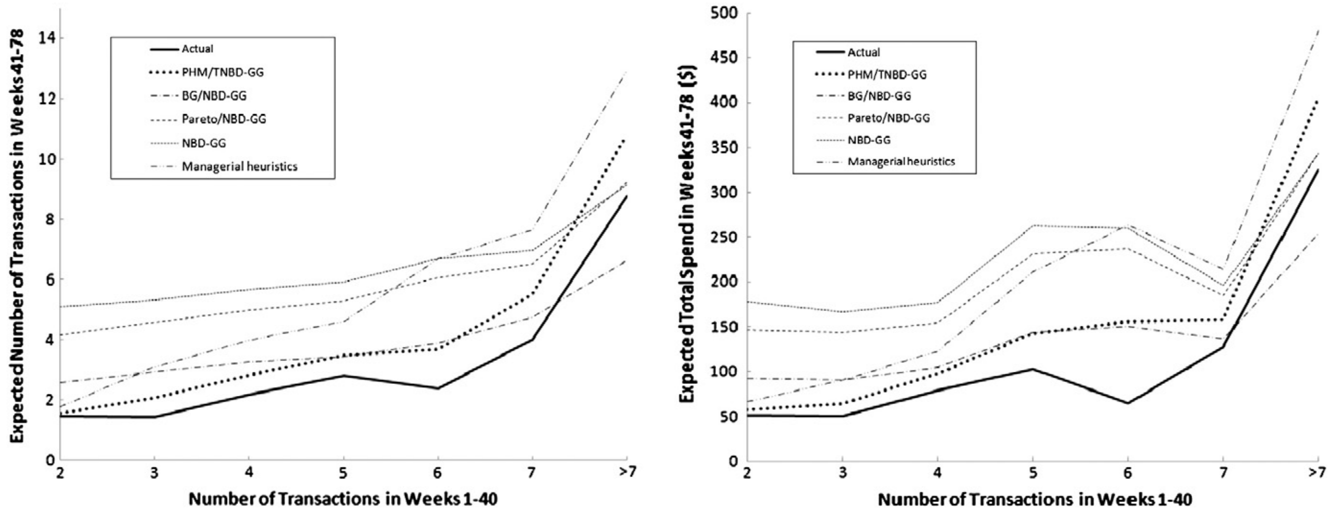

Fig. 5. Conditional expectations of purchasing for dynamic customers of CDNOW.

modeling purchase frequency and average monetary value. Furthermore, our model allows customer behavior to change over time, customers to reestablish their relationship with the company after some periods of inactivity, and it relaxes the independence assumption between purchase frequency and monetary value. Application of our model to two datasets from two different industries shows that our model outperforms traditional stochastic models and managerial heuristics on individual customer measures, relevant for CLV computation, i.e. purchase incidence, frequency and total expenditure. Model performance especially improved for those customers who show dynamic patterns of purchase behavior and when purchase frequency and monetary value are related. At the aggregate level, however, the traditional stochastic models (BG/NBD-GG and Pareto/NBD-GG) outperformed our proposed model. In follow-up analyses in which we selected customers with more dynamic purchase patterns, as defined by a negative autocorrelation of total expenditure over time, the PHM/TNBD-GG systematically outperforms all benchmarks at the individual customer level as well as at the aggregate level. In contrast, for customers with more stable purchase patterns, our model performs as the second best at the individual level and outperforms benchmarks at the aggregate level. These results indicate that our model provides adequate forecasts even in extreme situations.

In addition to more accurate predictions, our model also provides managerially valuable insights in actual customer behavior that previous models do not allow. For instance, the activity states represent different buying patterns. These patterns and the information about the switching probabilities from the estimated switching matrix are useful for marketing managers to target specific groups of customers based on their behavior and provides information about the dynamics in the market. In addition, in the hypermarket application, some states represent a negative relationship between purchase frequency and monetary value. This illustrates that although customers may have a similar total spending across periods, their behavior may be very different; some customers visiting very frequently and buying

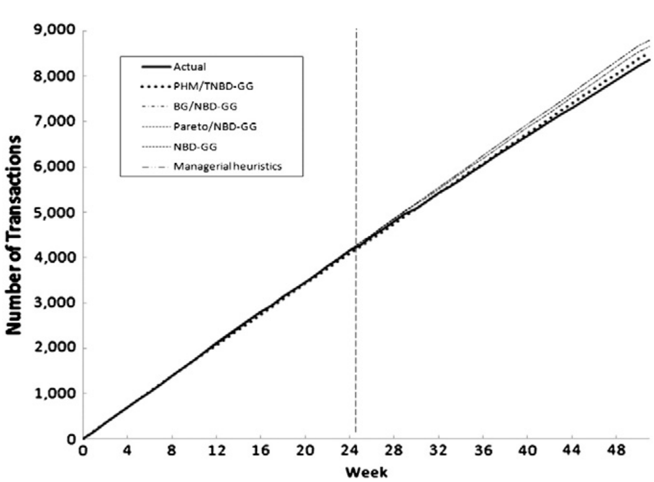

Fig. 6. Weekly time-series tracking plot for dynamic customers of hypermarket.

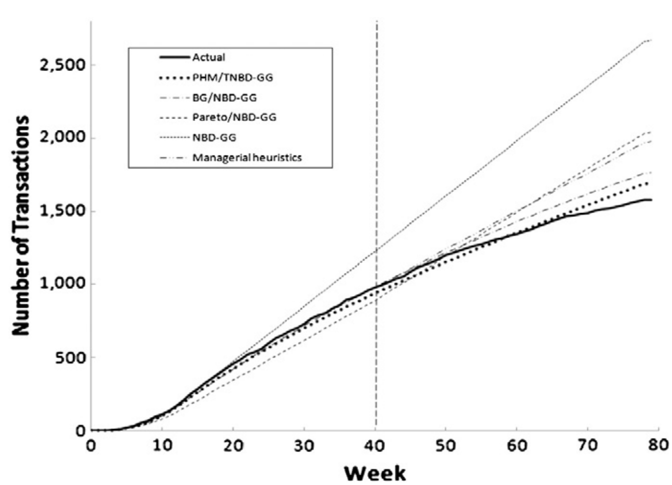

Fig. 7. Weekly time-series tracking plot for dynamic customers of CDNOW.

Table 10
Out-of-sample forecasts for static customers of hypermarket.

Criterion	PHM/TNBD-GG	BG/NBD-GG	Pareto/NBD-GG	NBD-GG	Managerial heuristics
Individual customers measures (*MAD*)					
Total expenditure	16.50 (.27)	16.69 (.29)	16.25 (.28)	21.25 (.24)	20.73 (.21)
Purchase frequency	.47 (.0063)	.48 (.0068)	.46 (.0066)	.60 (.0054)	.59 (.0057)
Purchase incidence	.30 (.0029)	.32 (.0028)	.32 (.0029)	.41 (.0019)	.38 (.0024)
Aggregate measures (*MAPE*)					
Total expenditure	18.28 (2.28)	28.36 (2.81)	28.40 (2.87)	37.85 (4.95)	35.89 (4.85)
Purchase frequency	12.72 (1.74)	27.07 (2.18)	26.62 (2.21)	40.23 (3.38)	38.61 (3.34)

Note: MAD = mean absolute deviation, MAPE = mean absolute percentage error, standard errors between brackets.

small amounts, others visiting infrequently and spending more. This may have different implications for customer profitability as the former may be more costly. Furthermore, model comparisons across data sets revealed that switching patterns differ across industries, which probably reflect when and how customers plan their purchases. Customers of the hypermarket have a much smaller probability to switch to inactivity after making a purchase. This implies that the hypermarket might need to plan their promotions for shorter time periods than CDNOW (e.g. weekly vs. monthly).

Although our model alleviates some of the assumptions of earlier models, it still possesses some limitations, which lead to interesting avenues for future research. First, we developed our model to deal with noncontractual business situations. In these situations, defection is unobserved. However, in various business settings, defection is observed, such as subscriptions to magazines, mobile telephone providers, or memberships to organizations (Ascarza and Hardie 2012; Borle, Singh and Jain (2008)). While existing stochastic models are not directly able to describe these settings, our PHM/TNBD-GG fits these situations well. Note that we explicitly model the defection state, which is in our empirical setting latent. However, by slightly modifying the supervision function in our partially hidden Markov estimation procedure, it is relatively straightforward to include observed defections. It would be interesting for future research to apply our model to such contractual settings. Second, although dynamic purchase patterns are common in situations such as variety seeking and new product introductions (Fader, Hardie, and Huang 2004), these purchase patterns may also be caused by marketing activities. For instance, customers may increase purchase frequency due to advertising campaigns or sales promotions. It would be fruitful to extend our model to incorporate such marketing activities. We suggest relating these marketing activities to the Markov switching matrix, such that the switching probabilities are a function of these activities (see also Netzer, Lattin, and Srinivasan 2008). Similarly, the incorporation of other explanatory variables in the Markov switching matrix could also improve model performance. For instance, purchase behavior may exhibit seasonal patterns. This seasonality could be taken into account through the inclusion of concomitant seasonal factors in the transition probabilities. A third direction for future research is to analyze the impact of the length of the aggregation period (weeks in this research) on model performance. Choosing a correct length that corresponds to customer planning horizons could affect model fit. Furthermore, it is likely that within industries these planning horizons vary across customers and thus also switching periods. Another possible extension of our model is to allow higher-order switching patterns that are violated by the first-order Markov assumption. Such higher-order patterns could capture whether customers who switch from activity to inactivity and back to activity are coming back to the same activity state.

In conclusion, we introduced a new stochastic model of customer behavior that deals with a broader setting of customer purchase patterns than existing stochastic models. While stochastic models have received much attention, some academics and practitioners are still skeptical about their performance (Wübben and Wangenheim 2008). We believe that our model, which alleviates some of the restrictive behavioral assumptions of previous models and in addition provides useful managerial insights, could be a powerful tool for CLV analysis.

Table 11
Out-of-sample forecasts for static customers of CDNOW.

Criterion	PHM/TNBD-GG	BG/NBD-GG	Pareto/NBD-GG	NBD-GG	Managerial heuristics
Individual customers measures (*MAD*)					
Total Expenditure	3.64 (.18)	3.40 (.18)	3.90 (.17)	7.39 (.21)	4.47 (.17)
Purchase frequency	.10 (.0036)	.09 (.0036)	.11 (.0036)	.19 (.0031)	.12 (.0034)
Purchase incidence	.09 (.0030)	.08 (.0030)	.09 (.0043)	.23 (.0021)	.11 (.0029)
Aggregate measures (*MAPE*)					
Total expenditure	44.86 (7.73)	48.02 (7.69)	45.92 (10.47)	225.34 (33.14)	70.02 (14.68)
Purchase frequency	33.23 (5.29)	42.58 (3.99)	32.60 (7.31)	180.63 (22.50)	48.55 (10.67)

Note: MAD = mean absolute deviation, MAPE = mean absolute percentage error, standard errors between brackets.

Appendix A. Derivation of Purchase Frequency for a Randomly Chosen Customer

We assume that if a customer i is in activity state k in time period t, purchase frequency follows a zero-truncated Poisson distribution with parameter λ_{ikt} (i.e. an active consumer purchases at least once). The truncated Poisson is defined as follows:

$$P(X=x|\lambda_{ikt}) = \frac{\lambda_k^x e^{-\lambda_k}}{x!(1-e^{-\lambda_k})} \quad \forall x=1,2,3,\ldots \tag{A1}$$

The purchase rate λ_{ikt} is not homogeneous for every customer that follows a purchase pattern k – it is distributed according to the modified gamma distribution with parameters r_k and α_k:

$$P(\lambda_{ikt}|r_k,\alpha_k) = \frac{{\alpha_k}^{r_k}(\alpha_k+1)^{r_k}}{\left((\alpha_k+1)^{r_k}-\alpha^{r_k}\right)\Gamma(r_k)}\lambda_{ikt}^{r_k-1}e^{-\alpha_k\lambda_{ikt}}\left(1-e^{-\lambda_{ikt}}\right) \quad \forall\lambda_{ikt}>0. \tag{A2}$$

The distribution of the number of purchases X for a randomly chosen customer in activity state k can be computed as follows:

$$P(X=x|r_k,\alpha_k) = \int_0^\infty P(X=x|\lambda_{ikt})\cdot P(\lambda_{ikt}|r_k,\alpha_k)\mathrm{d}\lambda_{ikt}. \tag{A3}$$

By taking into account the following identity: $\int_0^\infty x^n e^{-ax}\mathrm{d}x = \frac{\Gamma(n+1)}{a^{n+1}}$, Eq. (A3) leads to a zero-truncated negative binomial distribution (A4) for the number of purchases for a randomly chosen consumer:

$$\begin{aligned} P(X=x|r_k,\alpha_k) &= \int_0^\infty \frac{\lambda_{ikt}^x e^{-\lambda_{ikt}}}{x!(1-e^{-\lambda_{ikt}})}\cdot\frac{{\alpha_k}^{r_k}(\alpha_k+1)^{r_k}}{\left((\alpha_k+1)^{r_k}-\alpha^{r_k}\right)\Gamma(r_k)}\lambda_{ikt}^{r_k-1}e^{-\alpha_k\lambda_{ikt}}\left(1-e^{-\lambda_{ikt}}\right)\mathrm{d}\lambda_{ikt} \\ &= \frac{{\alpha_k}^{r_k}(\alpha_k+1)^{r_k}}{\left((\alpha_k+1)^{r_k}-\alpha^{r_k}\right)\Gamma(r_k)x!}\int_0^\infty \lambda_{ikt}^{x+r_k-1}e^{-\lambda_{ikt}(1+\alpha_k)}\mathrm{d}\lambda_k \\ &= \frac{{\alpha_k}^{r_k}(\alpha_k+1)^{r_k}}{\left((\alpha_k+1)^{r_k}-\alpha^{r_k}\right)\Gamma(r_k)x!}\cdot\frac{\Gamma(x+r_k)}{(\alpha_k+1)^{x+r_k}} \\ &= \frac{\Gamma(x+r_k)}{\Gamma(r_k)x!}\cdot\frac{{\alpha_k}^{r_k}}{(\alpha_k+1)^{r_k}-\alpha^{r_k}}\cdot\frac{1}{(\alpha_k+1)^x}. \end{aligned} \tag{A4}$$

with expectation: $\frac{(\alpha_k+1)^{r_k}}{(\alpha_k+1)^{r_k}-\alpha_k^{r_k}}\cdot\frac{r_k}{\alpha_k}$.

Appendix B. Model Estimation

We estimated our model parameters using the expectation-maximization (EM) algorithm (Wedel and Kamakura 2000). The EM algorithm is particularly useful in cases where some parts of the likelihood function to be maximized cannot be observed. The EM builds an auxiliary function assuming some initial values for the model parameters (expectation step). This auxiliary function, after some simplification, allows finding an estimate of the model parameters using standard maximization techniques (maximization step). This new set of parameters is then used as new values in the expectation step and the process starts again. The algorithm stops when a convergence criterion has been reached.

In order to apply the EM algorithm to our problem we built upon its version for hidden Markov models estimation (see Rabiner 1989 for a detailed tutorial), and included two modifications. First, due to our treatment of inactivity and defection as zero purchases that are observed, we followed Zhang and Mason (1989) and modified the probabilities of being in inactivity, defection or activity states. Second, because the EM algorithm for hidden Markov model estimation was originally developed for speech recognition systems in which data input is a long series of observations from one subject, we followed Li, Parizeau, and Plamondon (2000) to apply the algorithm to relatively short series of many individuals. In particular, we apply the EM algorithm at an individual level, for every subject, and aggregate the results by averaging the individual parameter estimates at each iteration.

Let $M = \left\{\varphi_{ik_1}, \phi_{k_{t-1}k_t}, r_k, \alpha_k, u_k, w_k, \delta_k\right\}$ be the set of model parameters, Q be the set of all possible state transitions, and $O_i = \{O_{i1} = (x_{i1},m_{i1}), O_{i2} = (x_{i2},m_{i2}),.., O_{iT} = (x_{iT},m_{iT})\}$ the set of observations consisting of number of transactions and monetary value of customer i in each time period t. In general, the auxiliary function of a hidden Markov model equals (Rabiner 1989):

$$F_i(M,\overline{M}) = \sum_Q P(Q|O_i,M)\log\left[P(O_i,Q|\overline{M})\right] \tag{B1}$$

where $\overline{M} = \left\{\overline{\varphi}_{ik_1}, \overline{\phi}_{k_{t-1}k_t}, \overline{r}_k, \overline{\alpha}_k, \overline{u}_k, \overline{w}_k, \overline{\delta}_k\right\}$ represents the re-estimated model parameters given M. As shown by Rabiner (1989), the likelihood of $\overline{M}$ is larger or equal to M. Following Ephraim and Merhav (2002) the auxiliary function (B1) can be written as the sum of three components:

$$F_i\left(M, \overline{M}\right) = \sum_{k=1}^{K+2} P(S_{i1} = k|O_i, M)\log(\overline{\varphi}_{ik}) + \sum_{j,k=1}^{K+2}\sum_{t=2}^{T} P\left(S_{i,t-1} = j, S_{it} = k|O_i, M\right)\log\left(\overline{\phi}_{jk}\right)$$
$$+\sum_{k=1}^{K+2}\sum_{t=2}^{T} P\left(S_{it} = k|O_i, M\right)\log P\left(O_{it}|\overline{r}_{k_t}, \overline{\alpha}_{k_t}, \overline{u}_{k_t}, \overline{w}_{k_t}, \overline{\delta}_{k_t}\right), \quad \text{(B2)}$$

with S_{it} the state of consumer i in period t. An advantage of Eq. (B2) is that each of the three components can be maximized independently. In order to estimate the model, we therefore use the following 2 steps iteratively.

Step 1: initial state probabilities (φ_{k_1}) *and switching probabilities* (ϕ_{k_{t-1},k_t})

To compute the initial state probabilities (φ_{k_1}) and the switching probabilities (ϕ_{k_{t-1},k_t}), we use Baum–Welch re-estimation formulas (see Rabiner 1989 for a detailed explanation), by taking into account restrictions on active, inactive and defected customers. The Baum–Welch re-estimation formulas take the following two formulas into account:

$$A_{it}(k) = P(O_{i1}, O_{i2}, .., O_{it}, k|M), \quad \text{(B3)}$$

$$B_{it}(k) = P\left(O_{i,t+1}, O_{i,t+2}, .., O_{iT}|k, M\right). \quad \text{(B4)}$$

In Eq. (B3), $A_{it}(k)$ represents the probability of observing $O_{i1}, O_{i2}, \ldots O_{it}$, until time t and being in state k at time t, given the model M. In Eq. (B4), $B_{it}(k)$ represents the probability of observing $O_{i,t+1}, O_{i,t+2}, \ldots, O_{iT}$ given being in state k in period t and the model parameters M. Both variables $A_{it}(k)$ and $B_{it}(k)$ are computed inductively for each customer i. For $A_{it}(k)$

a) *Initialization:*

$$A_{i1}(k) = \phi_{ik} P(x_{i1}, m_{i1}|k)$$

b) *Induction*:

$$A_{it}(k) = \left[\sum_{j=1}^{K+2} A_{i,t-1}(j)\varphi_{jk}\right] P(x_{it}, m_{it}|k),$$

Supervision function: If customer i makes zero purchases in period t:

$$A_{it}(k) = 0 \ \text{ for } \ k = 1, 2..K.$$

Supervision function: If customer i is active in period t:

$$A_{it}(k) = 0 \ \text{ for } \ k = K + 1 \ \text{ and } \ K + 2.$$

For $B_{it}(k)$:

a) *Initialization*, Supervision function: if customer i is active in period T:

$$B_{iT}(k) = 1 \ \text{ for } \ k = 1, 2..K, \ \text{ zero otherwise},$$

Supervision function: If customer i makes zero purchases in period T:

$$B_{iT}(k) = 1 \ \text{ for } \ k = K + 1 \ \text{ and } \ K + 2, \ \text{ zero otherwise}$$

b) *Induction*:

$$B_{it}(k) = \sum_{j=1}^{K+2} \varphi_{kj} P(x_{i,t+1}, m_{i,t+1}|k) B_{i,t+1}(k)$$

Supervision function: if customer i makes zero purchases in period t:

$$B_{it}(k) = 0 \text{ for } k = 1, 2, \ldots, \mathrm{K},$$

Supervision function: if customer i is active in period t:

$$B_{it}(k) = 0 \text{ for } k = K+1 \text{ and } \mathrm{K}+2$$

Using $A_{it}(k)$ and $B_{it}(k)$, we can calculate for each individual i:

- the probability of being in state k at time t given the observation sequence O_i and the model M:

$$G_{it}(k) = P(k|O_i, M) = \frac{A_{it}(k) B_{it}(k)}{\sum_{j=1}^{K+2} A_{it}(j) B_{ij}(j)} = E(p_{ikt}). \tag{B5}$$

- the probability of being in state j at time t-1 and in state k at time t, given the observation sequence O_i and the model M:

$$H_{it}(j,k) = \frac{A_{i,t-1}(j) \varphi_{jk} P(x_{it}, m_{it}|k) B_{it}(k)}{\sum_{r=1}^{K+2} \sum_{s=1}^{K+2} A_{i,t-1}(r) \varphi_{rs} P(x_{it}, m_{it}|s) B_{it}(s)} \text{ for all } j \text{ and } k \in \{1, .., K+2\}. \tag{B6}$$

Using previous expressions, we are able to compute the initial probabilities $\overline{\phi}_{ik}$ and switching probabilities $\overline{\varphi}_{jk}$ as follows:
For all active customers in period 1:

$$\overline{\phi}_k = \frac{1}{N_{\text{active in period 1}}} \sum_{i \in \text{Active in period 1}}^{N_{\text{active in period 1}}} G_{i1}(k) \text{ for } k = 1, 2, \ldots K, \text{ zero otherwise}, \tag{B7}$$

where $N_{\text{active in period 1}}$ is the number of active customers in the customer base in period 1. For customers who make zero purchases in period 1:

$$\overline{\phi}_k = 1 \text{ for } k_1 = K+1, \text{ zero otherwise}. \tag{B8}$$

Since all the customers are required to be active or inactive at the beginning of the observation period, i.e. not defected, $\varphi_{K+2} = 0$. For a specific customer i, the expected switching probability $E(r_{ijk})$ from state j to state k, independent of time period t equals:

$$E(r_{ijk}) = \sum_{t=2}^{T} H_{it}(j,k) / \sum_{t=2}^{T} G_{i,t-1}(j), \text{ for all } j \text{ and } k \in \{1, .., K+2\}. \tag{B9}$$

For both active and inactive customers, the average switching probabilities from state j to k over all customers are computed as follows:

$$\overline{\varphi}_{jk} = \sum_{i=1}^{N} \sum_{t=2}^{T} H_{it}(j,k) / \sum_{i=1}^{N} \sum_{t=2}^{T} G_{i,t-1}(j), \text{ for all } j \text{ and } k \in \{1, .., K+2\}. \tag{B10}$$

Because switching to permanent defection $E(r_{ij,K+2})$ for customer i from state $j \in \{1,.., K+1\}$ equals zero if this customer makes a purchase in the final period T, we use a weighted average of $E(r_{ijk})$ and $\overline{\varphi}_{jk}$ as our final estimate of the individual specific switching matrix, which is computed as follows:

$$E\left(q_{ijk}\right) = \frac{\sum_{t=1}^{T} G_{it}(j) \cdot E(r_{ijk}) + \overline{\varphi}_{jk}}{1 + \sum_{t=1}^{T} G_{it}(j)}, \text{ for all } j \text{ and } k \in \{1, .., K+2\}. \tag{B11}$$

Step 2: estimation of $\overline{r}_k, \overline{\alpha}_k, \overline{u}_k, \overline{w}_k, \overline{\delta}_k$

In order to estimate $\overline{r}_k, \overline{\alpha}_k, \overline{u}_k, \overline{w}_k, \overline{\delta}_k$, we only need to maximize the third part of the auxiliary function (B2) across all customers:

$$\sum_{i=1}^{N}\sum_{k=1}^{K+2}\sum_{t=2}^{T} P(S_{it}=k|O_i,M)\log P\left(O_{it}|\overline{r}_k,\overline{\alpha}_k,\overline{u}_k,\overline{w}_k,\overline{\delta}_k\right). \tag{B12}$$

Note that optimizing Eq. (B12) is relatively straightforward, using standard maximum likelihood optimization, as $P(O_{it}|\overline{r}_k, \overline{\alpha}_k, \overline{u}_k, \overline{w}_k, \overline{\delta}_k) = P\left(x_{it}|\overline{r}_k, \overline{\alpha}_k\right)\cdot P\left(m_{it}|\overline{u}_k, \overline{w}_k, \overline{\delta}_k\right)$, which is the product of the zero-truncated NBD for purchase frequency (see Appendix A) and the gamma–gamma distribution (see Colombo and Jiang 1999).

After obtaining all model parameters $\overline{M} = \left\{\overline{\varphi}_{ik_1}, \overline{\phi}_{k_{t-1}k_t}, \overline{r}_k, \overline{\alpha}_k, \overline{u}_k, \overline{w}_k, \overline{\delta}_k\right\}$ in steps 1 and 2, we use them as initial values of the algorithm and re-run steps 1 and 2 until $F_i\left(\overline{M}, M\right) = F_i(M, M)$ (Ephraim and Merhav 2002).

Appendix C. Derivation of Expected Number of Purchases in State *k*

In this Appendix we derive the expected number of purchases, given that a customer is in state k. In our derivation we use Hurwitz zeta function, which is defined as:

$$\zeta(s,a) = \frac{1}{\Gamma(s)}\int_0^\infty \frac{t^{s-1}}{e^{at}(1-e^{-t})}dt. \tag{C1}$$

Similar to Euler's integral for the Gaussian hypergeometric function, which is central in the derivation of the BG/NBD model (Fader, Hardie, and Lee 2005a), evaluation for a given set of parameters is relatively straightforward (even in programs such as Excel) as it can be approximated with a polynomial series:

$$\zeta(s,a) = \sum_{k=0}^{\infty}\frac{1}{(k+a)^s}. \tag{C2}$$

Given a customer i being in activity state k in period t, purchase frequency of this customer follows a truncated Poisson distribution with parameter λ_{ikt}:

$$P(X_{ik}=x|\lambda_{ikt}) = \frac{(\lambda_{ikt})^x e^{-\lambda_{ikt}}}{x!(1-e^{-\lambda_{ikt}})} \tag{C3}$$

Given parameters r_k and α_k and the assumption that λ_{ikt} has an alternative gamma distribution, the likelihood of this customer's purchase frequency is (see Appendix A):

$$P(X_{ik}=x|r_k,\alpha_k) = \frac{\Gamma(x+r_k)}{\Gamma(r_k)x!}\cdot\frac{{\alpha_k}^{r_k}}{(\alpha_k+1)^{r_k}-\alpha^{r_k}}\cdot\frac{1}{(\alpha_k+1)^x} \tag{C4}$$

Using Bayes theorem, and the assumption that the number of purchases follows a truncated Poisson distribution, the expected number of purchases in state k for a random customer making x_{it} purchases $E(X_{ik}|r_k,\alpha_k,x_{it}) = E\left(\frac{\lambda_{ikt}}{1-\exp(-\lambda_{ikt})}|r_k,\alpha_k,x_{it}\right)$ equals:

$$\begin{aligned} E\left(\frac{\lambda_{ikt}}{1-e^{-\lambda_{ikt}}}|r_k,\alpha_k,x_{it}\right) &= \int_0^\infty \frac{\frac{\lambda_{ikt}}{1-e^{-\lambda_{ikt}}}P(X_{ik}=x_{it}|\lambda_{ikt})\cdot P(\lambda_{ikt}|r_k,\alpha_k)}{P(X_{ik}=x_{it}|r_k,\alpha_k)}\mathrm{d}\lambda_{ikt} \\ &= \frac{(\alpha_k+1)^{r_k+x_{it}}}{\Gamma(x_{it}+r_k)}\cdot\int_0^\infty\frac{{\lambda_{ikt}}^{x_{it}+r_k}e^{-\lambda_{ikt}(\alpha_k+1)}}{(1-e^{-\lambda_{ikt}})}\mathrm{d}\lambda_{ikt}. \end{aligned} \tag{C5}$$

Recalling Hurwitz zeta function in Eq. (C1), Eq. (C5) can be expressed as follows:

$$E\left(\frac{\lambda_{ikt}}{1-e^{-\lambda_{ikt}}}|r_k,\alpha_k,x_{it}\right) = (\alpha_k+1)^{r_k+x_{it}}\cdot(x_{it}+r_k)\cdot\zeta(x_{it}+r_k+1,\alpha_k+1) \tag{C6}$$

Appendix D. Forecasting Purchase Incidence in the BG/NBD Model

Following Fader, Hardie, and Lee (2005a, equation 3, p.277), the individual-level likelihood for the BG/NBD model is[4]:

$$L\left(x, t_x|\lambda, p, T\right) = (1-p)^x \lambda^x e^{-\lambda T} + p(1-p)^{x-1} \lambda^x e^{-\lambda t_x}. \tag{D1}$$

The first part in this equation contains the likelihood of being alive at time T:

$$L(\text{alive at } T|\lambda, p, x, t_x) = (1-p)^x \lambda^x e^{-\lambda T}. \tag{D2}$$

Furthermore, given the likelihood that a customer is alive at time T, the likelihood that this customer is still alive at $T + t_a$ equals Eq. (D2) multiplied by the likelihood that a customer does not defect in the time period T to $T + t_a$. This likelihood is given by Fader, Hardie, and Lee (2005a, p.278), and equals $e^{-\lambda p t_a}$. Given that a customer is alive at time period $T + t_a$ the likelihood of purchase incidence in period $T + t_a$ to $T + t_b$ equals $1-e^{-\lambda(t_b-t_a)}$. Hence, the likelihood of purchase incidence in period $T + t_a$ to $T + t_b$ equals:

$$L(t_a, t_b, x|\lambda, p, T) = (1-p)^x \lambda^x e^{-\lambda(T+pt_a)}\left(1-e^{-\lambda(t_b-t_a)}\right). \tag{D3}$$

Given that λ follows a gamma distribution and p a beta distribution, and applying Bayes' rule, the probability of purchase incidence given the purchase history of a customer equals:

$$P(Incidence(t_a, t_b)|r, \alpha, a, b, T, t_x, x) = \int_0^1\int_0^\infty \frac{(1-p)^x \lambda^x e^{-\lambda(T+pt_a)}\left(1-e^{-\lambda(t_b-t_a)}\right)\cdot P(\lambda|r,\alpha)\cdot P(p|a,b)}{P(x, t_x|T, r, \alpha, a, b)} d\lambda dp. \tag{D4}$$

Fader, Hardie, and Lee (2005a, equation (6), p.278) provide the equation of $P(x, t_x|T, r, \alpha, a, b)$, which equals:

$$P(x, t_x|T, r, \alpha, a, b) = \frac{B(a, b+x)}{B(a,b)} \cdot \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+T)^{r+x}} + \frac{B(a+1, b+x-1)}{B(a,b)} \cdot \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+t_x)^{r+x}}. \tag{D5}$$

Integrating out first λ in Eq. (D4) gives us the following:

$$\begin{aligned}
&\frac{(1-p)^x P(p|a,b)}{P(x,t_x|T,r,\alpha,a,b)} \int_0^\infty \lambda^x e^{-\lambda(T+pt_a)}\left(1-e^{-\lambda(t_b-t_a)}\right)\cdot P(\lambda|r,\alpha) d\lambda = \\
&= \frac{(1-p)^x P(p|a,b)}{P(x,t_x|T,r,\alpha,a,b)} \int_0^\infty \lambda^x e^{-\lambda(T+pt_a)}\left(1-e^{-\lambda(t_b-t_a)}\right)\cdot \frac{\alpha^r \lambda^{r-1} e^{-\lambda a}}{\Gamma(r)} d\lambda \\
&= \frac{\alpha^r (1-p)^x P(p|a,b)}{\Gamma(r) P(x,t_x|T,r,\alpha,a,b)} \int_0^\infty \lambda^{x+r-1} e^{-\lambda(T+pt_a+\alpha)}\left(1-e^{-\lambda(t_b-t_a)}\right) d\lambda \\
&= \frac{\alpha^r (1-p)^x P(p|a,b)}{\Gamma(r) P(x,t_x|T,r,\alpha,a,b)} \left(\int_0^\infty \lambda^{x+r-1} e^{-\lambda(T+pt_a+\alpha)} d\lambda - \int_0^\infty \lambda^{x+r-1} e^{-\lambda(T+pt_a+\alpha+t_b-t_a)} d\lambda\right) \\
&= \frac{\alpha^r (1-p)^x P(p|a,b)}{\Gamma(r) P(x,t_x|T,r,\alpha,a,b)} \left(\frac{\Gamma(x+r)}{(T+pt_a+\alpha)^{x+r}} - \frac{\Gamma(x+r)}{(T+pt_a+\alpha+t_b-t_a)^{x+r}}\right)
\end{aligned} \tag{D6}$$

Integrating Eq. (D6) over p gives us the probability of purchase incidence:

$$\begin{aligned}
&P(Incidence(t_a, t_b)|r, \alpha, a, b, T, t_x, x) = \\
&= \int_0^1 \frac{\alpha^r (1-p)^x P(p|a,b)}{\Gamma(r) P(x,t_x|T,r,\alpha,a,b)} \left(\frac{\Gamma(x+r)}{(T+pt_a+\alpha)^{x+r}} - \frac{\Gamma(x+r)}{(T+pt_a+\alpha+t_b-t_a)^{x+r}}\right) dp \\
&= \frac{\alpha^r \Gamma(x+r)}{\Gamma(r) P(x,t_x|T,r,\alpha,a,b)} \int_0^1 \left(\frac{(1-p)^x P(p|a,b)}{(T+pt_a+\alpha)^{x+r}} - \frac{(1-p)^x P(p|a,b)}{(T+pt_a+\alpha+t_b-t_a)^{x+r}}\right) dp \\
&= \frac{\alpha^r \Gamma(x+r)}{\Gamma(r) P(x,t_x|T,r,\alpha,a,b) B(a,b)} \int_0^1 \left(\frac{(1-p)^{x+b-1} p^{a-1}}{(T+pt_a+\alpha)^{x+r}} - \frac{(1-p)^{x+b-1} p^{a-1}}{(T+pt_a+\alpha+t_b-t_a)^{x+r}}\right) dp.
\end{aligned} \tag{D7}$$

[4] Note that we omit $\delta_{x>0}$ in the likelihood since we selected customers that made at least one purchase in the estimation period.

Note that by letting $q = 1 - p$ which implies that $dp = -dq$ (see Fader, Hardie, and Lee 2005a, p. 284),

$$\int_0^1 \left(\frac{(1-p)^{x+b-1} p^{a-1}}{(T + pt_a + \alpha)^{x+r}} \right) dp = \frac{1}{(\alpha + T + t_a)^{x+r}} \int_0^1 \left(\frac{q^{x+b-1}(1-q)^{a-1}}{\left(1 - \frac{t_a}{\alpha + T + t_a} q\right)^{x+r}} \right) dq = \frac{B(x+b,a)}{(\alpha + T + t_a)^{x+r}} {}_2F_1\left(r+x, x+b; a+x+b; \frac{t_a}{\alpha + T + t_a}\right), \tag{D8}$$

and

$$\int_0^1 \left(\frac{(1-p)^{x+b-1} p^{a-1}}{(T + pt_a + \alpha + t_b - t_a)^{x+r}} \right) dp = \frac{1}{(\alpha + T + t_b)^{x+r}} \int_0^1 \left(\frac{q^{x+b-1}(1-q)^{a-1}}{\left(1 - \frac{t_a}{\alpha + T + t_b} q\right)^{x+r}} \right) dq = \frac{B(x+b,a)}{(\alpha + T + t_b)^{x+r}} {}_2F_1\left(r+x, x+b; a+x+b; \frac{t_a}{\alpha + T + t_b}\right). \tag{D9}$$

With ${}_2F_1(a,b;c;z) = \frac{1}{B(b,c-b)} \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-zt)^a} dt$ Euler's integral for the Gaussian hypergeometric function. Substituting Eqs. (D9) and (D10) into Eq. (D8) gives us the probability of purchase incidence:

$$P(Incidence(t_a,t_b)|r,\alpha,a,b,T,t_x,x) = \frac{\alpha^r \Gamma(x+r) B(x+b,a)}{\Gamma(r) P(x,t_x|T,r,\alpha,a,b) B(a,b)} \left\{ \frac{{}_2F_1\left(r+x, x+b; a+x+b; \frac{t_a}{\alpha+T+t_a}\right)}{(\alpha+T+t_a)^{x+r}} - \frac{{}_2F_1\left(r+x, x+b; a+x+b; \frac{t_a}{\alpha+T+t_b}\right)}{(\alpha+T+t_b)^{x+r}} \right\}. \tag{D10}$$

Appendix E. Forecasting Purchase Incidence in the Pareto/NBD Model

For the derivation of purchase incidence in the Pareto/NBD model, we will need to evaluate an integral of the following form:

$$C(r,\alpha,s,\beta;\gamma,\delta,\varepsilon,\phi,\varphi) = \int_0^\infty \int_0^\infty \frac{\lambda^\gamma \mu^\delta}{(\lambda+\mu)^\phi} \cdot e^{-(\lambda+\mu)\varphi - \varepsilon\mu} g(\lambda|r,\alpha) g(\mu|s,\beta) \mathrm{d}\lambda \mathrm{d}\mu. \tag{E1}$$

Note that this double integral is similar to the double integral A by Fader and Hardie (2005, equation (7) p. 3), except for the terms $\phi \in \{0,1\}$ and $\varphi > 0$, and the term $\varepsilon\mu$ in the exponent, with $\varepsilon > 0$. Using the transformation $\mu = (1 - p)z$ and $\lambda = pz$, similar to Fader and Hardie (2005), one can show that:

$$C(r,\alpha,s,\beta;\gamma,\delta,\varepsilon,\phi,\varphi) = \frac{\alpha^r \beta^s}{(\beta+\varphi+\varepsilon)^{r+s+\gamma+\delta-\phi}} \frac{\Gamma(r+\gamma)\Gamma(s+\delta)}{\Gamma(r)\Gamma(s)} \left(\frac{1}{r+s+\gamma+\delta-1}\right)^\phi \times {}_2F_1\left(r+s+\gamma+\delta-\phi, r+\gamma; r+s+\gamma+\delta; \frac{\beta+\varepsilon-\alpha}{\beta+\varepsilon+\varphi}\right). \tag{E2}$$

Note that Eq. (E2) converges if $\alpha \le \varphi + 2\beta + 2\varepsilon$. If this inequality does not hold, we apply the transformation $\mu = pz$ and $\lambda = (1 - p)z$, which gives the following solution to Eq. (E1):

$$C(r,\alpha,s,\beta;\gamma,\delta,\varepsilon,\phi,\varphi) = \frac{\alpha^r \beta^s}{(\alpha+\varphi)^{r+s+\gamma+\delta-\phi}} \frac{\Gamma(r+\gamma)\Gamma(s+\delta)}{\Gamma(r)\Gamma(s)} \left(\frac{1}{r+s+\gamma+\delta-1}\right)^\phi \times {}_2F_1\left(r+s+\gamma+\delta-\phi, s+\delta; r+s+\gamma+\delta; \frac{\alpha-\varepsilon-\beta}{\alpha+\varphi}\right). \tag{E3}$$

Note that if $\alpha > \varphi + 2\beta + 2\varepsilon$, $\left|\frac{\alpha-\varepsilon-\beta}{\alpha+\varphi}\right|<1$, and Eq. (E3) converges. Following Fader and Hardie (2005, equation (31) p.12), the individual-level likelihood for being alive at time T is:

$$L(\text{alive at } T|\lambda,\mu,x,t_x) = \frac{\lambda^x e^{-(\lambda+\mu)T}}{L(x,t_x,T|\lambda,\mu).} \tag{E4}$$

Given the likelihood that a customer is alive at time T, the likelihood that this customer is still alive at $T + t_a$ equals Eq. (E4) multiplied by the likelihood that a customer does not defect in the time period T to $T + t_a$. Because a customer dies at rate μ, this likelihood equals $e^{-\mu t_a}$. Given that a customer is alive at time period $T + t_a$, there are two situations in which a customer can make a purchase in period $T + t_a$ to $T + t_b$. First, with probability $e^{-\mu(t_b-t_a)}$ that this customer is still alive at time t_b, the likelihood of purchase incidence in period $T + t_a$ to $T + t_b$ equals $1-e^{-\lambda(t_b-t_a)}$. Second, with probability $\mu e^{-\mu(\tau-t_a)}$ that this customer dies at time τ in period $T + t_a$ to $T + t_b$, the probability of purchase incidence in period $T + t_a$ to $T + \tau$ equals $1-e^{-\lambda(\tau-t_a)}$. Integrating over τ results in the following likelihood in this situation: $\int_{t_a}^{t_b}\left(1-e^{-\lambda(\tau-t_a)}\right)\mu e^{-\mu(\tau-t_a)}d\tau = \left(1-e^{-\mu(t_b-t_a)}\right)-\frac{\mu}{\lambda+\mu}\left(1-e^{-(\lambda+\mu)(t_b-t_a)}\right)$. Hence, the likelihood of purchase incidence in period $T + t_a$ to $T + t_b$ equals:

$$L(t_a,t_b|\lambda,p,T,t_x,x) = \frac{\lambda^x e^{-(\lambda+\mu)T}\left\{e^{-\mu t_b}\left(1-e^{-\lambda(t_b-t_a)}\right)+e^{-\mu t_a}\left(\left(1-e^{-\mu(t_b-t_a)}\right)-\frac{\mu}{\lambda+\mu}\left(1-e^{-(\lambda+\mu)(t_b-t_a)}\right)\right)\right\}}{L(x,t_x,T|\lambda,\mu).} \tag{E5}$$

Using Eq. (E5) and taking expectations over the joint distribution of λ and μ, which is updated with the information (x,t_x,T), we get:

$$P(Incidence(t_a,t_b)|r,\alpha,s,\beta,T,t_x,x) = \int_0^\infty\int_0^\infty L(t_a,t_b|\lambda,p,T,t_x,x)\cdot g(\lambda,\mu|r,\alpha,s,\beta,x,t_x,T)d\lambda d\mu. \tag{E6}$$

Fader, Hardie, and Lee (2005b, equation (A1), p.428) provide the equation of $P(x, t_x|T, r, \alpha, s, \beta)$, which equals:

$$P(x,t_x|T,r,\alpha,s,\beta) = \frac{\Gamma(r+x)\alpha^r\beta^s}{\Gamma(r)}\left\{\frac{1}{(\alpha+T)^{r+x}(\beta+T)^s}+\left(\frac{s}{r+s+x}\right)A_0\right\}, \tag{E7}$$

where, for $\alpha\geq\beta$ $A_0 = \frac{{}_2F_1\left(r+s+x,s+1;r+s+x+1;\frac{\alpha-\beta}{\alpha+t_x}\right)}{(\alpha+t_x)^{r+s+x}}-\frac{{}_2F_1\left(r+s+x,s+1;r+s+x+1;\frac{\alpha-\beta}{\alpha+T}\right)}{(\alpha+T)^{r+s+x}}$, and for $\alpha<\beta$ $A_0 = \frac{{}_2F_1\left(r+s+x,r+x;r+s+x+1;\frac{\beta-\alpha}{\beta+t_x}\right)}{(\beta+t_x)^{r+s+x}}-\frac{{}_2F_1\left(r+s+x,r+x;r+s+x+1;\frac{\beta-\alpha}{\beta+T}\right)}{(\beta+T)^{r+s+x}}$.

Applying Bayes' rule to the joint posterior of λ and μ in Eq. (E6), we get (see Fader and Hardie 2005, equation (33), p.12):

$$g(\lambda,\mu|r,\alpha,s,\beta,x,t_x,T) = \frac{L(x,t_x,T|\lambda,\mu)g(\lambda|r,\alpha)g(\mu|s,\beta)}{P(x,t_x,T|r,\alpha,s,\beta)}. \tag{E8}$$

Substituting Eqs. (E8) and (E5) in Eq. (E6), we get

$$\begin{aligned} &P(Incidence(t_a,t_b)|r,\alpha,s,\beta,T,t_x,x) = \\ &= \frac{\int_0^\infty\int_0^\infty\left[\left\{e^{-\mu t_b}\left(1-e^{-\lambda(t_b-t_a)}\right)+e^{-\mu t_a}\left(\left(1-e^{-\mu(t_b-t_a)}\right)-\frac{\mu}{\lambda+\mu}\left(1-e^{-(\lambda+\mu)(t_b-t_a)}\right)\right)\right\}\lambda^x e^{-(\lambda+\mu)T}g(\lambda|r,\alpha)g(\mu|s,\beta)d\lambda d\mu\right]}{P(x,t_x,T|r,\alpha,s,\beta)} \\ &= \frac{\int_0^\infty\int_0^\infty\left[\lambda^x\left\{-e^{-(\lambda+\mu)(T+t_b-t_a)-\mu t_a}+e^{-(\lambda+\mu)T-\mu t_a}-\frac{\mu}{\lambda+\mu}e^{-(\lambda+\mu)T-\mu t_a}+\frac{\mu}{\lambda+\mu}e^{-(\lambda+\mu)(T+t_b-t_a)-\mu t_a}\right\}g(\lambda|r,\alpha)g(\mu|s,\beta)d\lambda d\mu\right]}{P(x,t_x,T|r,\alpha,s,\beta)}. \end{aligned} \tag{E9}$$

Note that the numerator in Eq. (E9) consists of the sum of four double integrals that are in the form of Eq. (E1). We compute these double integrals next:

$$B_1 = \int_0^\infty\int_0^\infty\lambda^x e^{-(\lambda+\mu)(T+t_b-t_a)-\mu t_a}g(\lambda|r,\alpha)g(\mu|s,\beta)d\lambda d\mu = C(r,\alpha,s,\beta;x,0,t_a,0,T+t_b-t_a), \tag{E10}$$

$$B_2 = \int_0^\infty \int_0^\infty \lambda^x e^{-(\lambda+\mu)T-\mu t_a} g(\lambda|r,\alpha) g(\mu|s,\beta) d\lambda d\mu = C(r,\alpha,s,\beta;x,0,t_a,0,T), \tag{E11}$$

$$B_3 = \int_0^\infty \int_0^\infty \frac{\lambda^x \mu}{\lambda+\mu} e^{-(\lambda+\mu)T-\mu t_a} g(\lambda|r,\alpha) g(\mu|s,\beta) d\lambda d\mu = C(r,\alpha,s,\beta;x,1,t_a,1,T), \tag{E12}$$

$$B_4 = \int_0^\infty \int_0^\infty \frac{\lambda^x \mu}{\lambda+\mu} e^{-(\lambda+\mu)(T+t_b-t_a)-\mu t_a} g(\lambda|r,\alpha) g(\mu|s,\beta) d\lambda d\mu = C(r,\alpha,s,\beta;x,1,t_a,1,T+t_b-t_a), \tag{E13}$$

Substituting terms (E7) and (E10) to (E13) in (E9), we get the following expression for purchase incidence for the Pareto/NBD model.

$$P(Incidence(t_a,t_b)|r,\alpha,s,\beta,T,t_x,x) = \frac{-B_1+B_2-B_3+B_4}{\frac{\Gamma(r+x)\alpha^r\beta^s}{\Gamma(r)}\left\{\frac{1}{(\alpha+T)^{r+x}(\beta+T)^s}+\left(\frac{s}{r+s+x}\right)A_0\right\}}. \tag{E14}$$

Appendix F. Forecasting Purchase Incidence in the NBD Model

In the NBD model, a customer never dies and purchases at a frequency of λ. Therefore, the probability of purchase incidence in a specific future period $[t_a,t_b]$ equals $1-e^{-\lambda(t_b-t_a)}$. Given that λ follows a gamma distribution, applying Bayes' theorem, we get the following probability for a customer with x in the estimation period $[0,T]$:

$$\begin{aligned} P(Incidence(t_a,t_b)|x,T,r,\alpha) &= \frac{\int_0^\infty \left(1-e^{-\lambda(t_b-t_a)}\right) p(x,T|\lambda) p(\lambda|r,\alpha) \mathrm{d}\lambda_{ikt}}{p(x,T|r,\alpha)} \\ &= \frac{\frac{T^x\alpha^r}{x!\Gamma(r)}\left(\int_0^\infty \lambda^{x+r-1} e^{-\lambda(\alpha+T)} \mathrm{d}\lambda_k - \int_0^\infty \lambda^{x+r-1} e^{-\lambda(\alpha+T+t_b-t_a)} \mathrm{d}\lambda_k\right)}{\frac{T^x\alpha^r\Gamma(x+r)}{\Gamma(r)(\alpha+T)^{x+r}}} \\ &= 1-\left(\frac{\alpha+T}{\alpha+T+t_b-t_a}\right)^{x+r}. \end{aligned} \tag{F1}$$

References

Abe, Makoto (2009), "'Counting Your Customers' One by One: A Hierarchical Bayes Extention to the Pareto/NBD Model," *Marketing Science*, 28, 3, 541–53.

Allenby, Greg M., Robert P. Leone, and Lichung Jen (1999), "A Dynamic Model of Purchase Timing with Application to Direct Marketing," *Journal of the American Statistical Association*, 94, 446, 365–74.

Ascarza, Eva and Bruce G.S. Hardie (2012), "A Joint Model of Usage and Churn in Contractual Settings" working paper.

Batislam, Emine Persentili, Meltem Denizel, and Alpay Filiztekin (2007), "Empirical Validation and Comparison of Models for Customer Base Analysis," *International Journal of Research in Marketing*, 24, 3, 201–9.

Bawa, Kapil (1990), "Modeling Inertia and Variety Seeking Tendencies in Brand Choice Behavior," *Marketing Science*, 9, 3, 263–78.

Berger, Paul D. and Nada I. Nasr (1998), "Customer Lifetime Value: Marketing Models and Applications," *Journal of Interactive Marketing*, 12, 1, 17–29.

Blattberg, Robert C., Byung-Do Kim, and Scott A. Neslin (2008), *Database Marketing: Analyzing and Managing Customers*. New York, NY: Springer.

Borle, Sharad, Siddharth S. Singh, and Dipak C. Jain (2008), "Customer Lifetime Value Measurement," *Management Science*, 54, 1, 100–12.

Bucklin, Randolph E., Sunil Gupta, and S. Siddarth (1998), "Determining Segmentation in Sales Response Across Consumer Purchase Behaviors," *Journal of Marketing Research*, 35, 2, 189–97.

Colombo, Richard and Weina Jiang (1999), "A Stochastic RFM Model," *Journal of Interactive Marketing*, 13, 3, 2–12.

Du, Rex Y. and Wagner A. Kamakura (2006), "Household Life Cycles and Lifestyles in the United States," *Journal of Marketing Research*, 43, 1, 121–32.

Ephraim, Yariv and Neri Merhav (2002), "Hidden Markov Processes," *IEEE Transactions on Information Theory*, 48, 6, 1518–69.

Fader, Peter S. and Bruce G.S. Hardie (2001), "Forecasting Repeat Sales at CDNOW: A Case Study," *Interfaces*, 31, 3, S94–S107.

———, ———, and Chun-Yao Huang (2004), "A Dynamic Changepoint Model for New Product Sales Forecasting," *Marketing Science*, 23, 1, 50–65.

——— and ——— (2005), "A Note on Deriving the Pareto/NBD Model and Related Expressions," accessed January 29, 2011, available at http://www.brucehardie.com/notes/009/2005.

———, ———, and Ka Lok Lee (2005a), "Counting your Customers the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Science*, 24, 2, 275–84.

———, ———, and ——— (2005b), "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research*, 42, 4, 415–30.

——— and ——— (2009), "Probability Models for Customer-Base Analysis," *Journal of Interactive Marketing*, 23, 61–9.

———, ———, and Jen Shang (2010), "Customer-Base Analysis in a Discrete-Time Noncontractual Setting," *Marketing Science*, 29, 6, 1086–108.

Ho, Teck-Hua, Young-Hoon Park, and Yong-Pin Zhou (2006), "Incorporating Satisfaction into Customer Value Analysis: Optimal Investment in Lifetime Value," *Marketing Science*, 25, 3, 260–77.

Hurwitz, Adolf (1882), "Einige Eigenschaften der Dirichlet'schen Functionen $F(s) = \sum \left(\frac{D}{n}\right) \cdot \frac{1}{n^s}$, die bei der Bestimmung der Classenanzahlen binärer quadratischer Formen auftreten," *Zeitschrift für Mathematik und Physik*, 27, 86–101.

Jerath, K., Peter S. Fader, and Bruce G.S. Hardie (2011), "New Perspectives on Customer 'Death' Using a Generalization of the Pareto/NBD Model," *Marketing Science*, 30, 5, 866–80.

Li, Xiaolin, Marc Parizeau, and Réjean Plamondon (2000), "Training Hidden Markov Models with Multiple Observations—A Combinatorial Method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 4, 371–7.

Ma, Shaohui and Joachim Büschken (2011), "Counting your Customers from an 'Always a Share' Perspective," *Marketing Letters*, 22, 243–57.

Netzer, Oded, James M. Lattin, and V. Srinivasan (2008), "A Hidden Markov Model of Customer Relationship Dynamics," *Marketing Science*, 27, 2, 185–204.

Park, Sungho and Sachin Gupta (2011), "A Regime-Switching Model of Cyclical Category Buying," *Marketing Science*, 30, 3, 469–80 ((Articles in Advance)).

Payne, Adrian and Pennie Frow (2005), "A Strategic Framework for Customer Relationship Management," *Journal of Marketing*, 69, 4, 167–76.

Pfeifer, Phillip.E. and Robert L. Carraway (2000), "Modeling Customer Relationships as Markov Chains," *Journal of Interactive Marketing*, 14, 2, 43–55.

Rabiner, Lawrence R. (1989), "A Tutorial of Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, 77, 2, 257–85.

Reinartz, Werner J. and V. Kumar (2000), "On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing," *Journal of Marketing*, 64, 4, 17–35.

——— and ——— (2003), "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing*, 67, 1, 77–99.

Rust, Roland T., Katherine N. Lemon, and Valarie A. Zeithaml (2004), "Return on Marketing: Using Customer Equity to Focus Marketing Strategy," *Journal of Marketing*, 68, 1, 109–27.

Scheffer, Tobias, Christian Decomain, and Stefan Wrobel (2001), "Active Hidden Markov Models for Information Extraction," "*Lecture Notes in Computer Science, Proceedings of the 4th International Conference on Advances in Intelligent Data Analysis*, 2189. 309–18.

Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), "Counting your Customers: Who Are They and What Will They Do Next?," *Management Science*, 33, 1, 1–24.

——— and Robert A. Peterson (1994), "Customer Base Analysis: An Industrial Purchase Process Application," *Marketing Science*, 13, 1, 41–67.

Schwartz, Eric M., Eric T. Bradlow, and Peter S. Fader (2012), "Model Selection Using Database Characteristics: Classification Methods and Application to the HMM and its Children", *working paper*. (http://ssrn.com/abstract=208567)).

Schweidel, David A., Eric T. Bradlow, and Peter S. Fader (2011), "Portfolio Dynamics for Customers Multiservice Provider," *Management Science*, 57, 3, 471–86.

Thompson, Craig S., Peter J. Thomson, and Xiagou Zheng (2007), "Fitting a Multisite Daily Rainfall Model to New Zealand Data," *Journal of Hydrology*, 340, 25–39.

Truyen, Tran The, Hung Hai Bui, Dinh Q. Phung, and Svetha Venkatesh (2008), "Learning Discriminative Sequence Models from Partially Labelled Data for Activity Recognition," in PRICAL, 5351.LNAI, T.-B. Ho and Z.-H. Zhou, editors. 903–12.

Venkatesan, Rajkumar and V. Kumar (2004), "A Customer Lifetime Value Framework for Customer Section and Resourced Allocation Strategy," *Journal of Marketing*, 68, 4, 106–25.

——— and Timothy Bohling (2007), "Optimal Customer Relationship Management Using Bayesian Decision Theory: An Application for Customers Selection," *Journal of Marketing Research*, 44, Nov, 579–94.

Wedel, Michel and Wagner A. Kamakura (2000), *Market Segmentation, Conceptual and Methodological Foundations*. Second edition. Boston: Kluwer.

Wübben, Markus and Florian V. Wangenheim (2008), "Instant Customer Base Analysis: Managerial Heuristics Often 'Get it Right'," *Journal of Marketing*, 72, 3, 82–93.

Zhang, X. and J.S. Mason (1989), "Improved Training Using Semi-hidden Markov Models in Speech Recognition," *ICASSP-89: 1989 International Conference on Acoustics, Speech and Signal Processing*, Glasgow (UK). 306–9.