

IBM Research Report

Wallet Estimation Models

**Saharon Rosset, Claudia Perlich, Bianca Zadrozny, Srujana Merugu,
Sholom Weiss, Rick Lawrence**

IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598



Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Wallet estimation models

Saharon Rosset, Claudia Perlich, Bianca Zadrozny,
Srujana Merugu, Sholom Weiss, Rick Lawrence
Predictive Modeling Group
IBM T. J. Watson Research Center
Yorktown Heights, NY

{srosset,reisz,zadrozny,smerugu,sholom,ricklawr}@us.ibm.com

Abstract

The wallet of a customer is defined as the total amount this customer can spend in a specific product category. This is a vital piece of information for planning marketing and sales efforts. We discuss the important problem of customer wallet estimation, while emphasizing the use of predictive modeling technologies to generate useful estimates, and minimizing reliance on primary research. We suggest several customer wallet definitions and corresponding evaluation approaches. Our main contribution is in presenting several new predictive modeling approaches which allow us to estimate customer wallets, despite the fact that these are typically unobserved. We present empirical results on the success of these modeling approaches, using a dataset of IBM customers.

1 Introduction

The total amount of money a customer can spend on a certain product category is a vital piece of information for planning and managing sales and marketing efforts. This amount is usually referred to as the customer's *wallet* (also called *opportunity*) for this product category. There are many possible uses for wallet estimates, including straightforward targeting of sales force and marketing actions towards large wallet customers and prospects. In a more sophisticated sales and marketing environment, the combination of classical propensity models for a particular product category with the knowledge of the wallet for the same category can direct the sales efforts: it allows a company to market not only to customers or potential customers with a large wallet, but also to those with a high probability of buying specific products in the category.

By combining the customer wallet estimates with the data on how much they spend with a particular seller, we can calculate the share-of-wallet that the seller

has of each customer for a given product category. This information allows the seller to target customers based on their growth potential, a combination of total wallet and share-of-wallet. The classical approach of targeting customers that have historically generated large amounts of revenue for the company (known as lifetime value modeling, see e.g., [9]) does not give enough importance to customers with a large wallet, but small share-of-wallet, which are the ones with presumably the highest potential for revenue growth.

Share-of-wallet is also important for detecting partial defection or silent attrition, which occurs when customers increase their spending in a given category, without increasing the amount purchased from a particular company [7]. In certain industries, customer wallets can be easily obtained from public data. For example, in the credit card industry, the card issuing companies can calculate the wallet size and respective share-of-wallet using credit records from the three major credit bureaus [2]. For most industries, however, no public wallet information is available at the customer level. In this case, there are two approaches used in practice for obtaining wallet estimates:

1. Top-Down: starts from a public aggregate estimate for the overall industry opportunity in a given country and splits this estimate across the individual customers using heuristics based on the customer characteristics. For example, if the customers are companies, the overall opportunity could be divided among the companies proportionally to their number of employees.
2. Bottom-Up: estimates the wallet directly at the customer level, using heuristics or predictive models based on customer information. A common approach is to obtain actual wallet information for a random subset of customers/prospects through primary research. A model is then developed based on this data to predict the wallet for the other customers/prospects.

Although customer wallet and share-of-wallet have been recognized as important customer value metrics in the marketing and services literature for a number of years [7, 8, 3, 4], there is not much published work on actual wallet modeling. The few references available are limited to commercial white papers from marketing analytics consulting companies [2] and a very recent thesis proposal [1].

The white paper from Epsilon [2] describes at a very high level the methodology that they use to estimate wallets for both customers and prospects at a given product category. Epsilon uses a bottom-up approach where a survey provides the self-reported category wallet for a sample of customers. The self-reported wallet is used to develop a multivariate linear regression that can be applied to the whole market or customer base. They do not describe which variables are used in the model and do not provide experimental results. In [1], the authors propose a Multivariate Latent Factor Model that a bank can use to impute a customers' holdings of financial products outside the bank, based on a combination of internally available data and survey information about customers' holdings with competitors.

In this paper, we address the issue of predicting the wallet for IBM customers that purchase Information Technology (IT) products such as servers, software, and services. Our models are bottom-up, with explanatory features drawn from historical transaction data as well as firmographic data taken from external sources like Dun & Bradstreet. We take a predictive modeling approach, and attempt to minimize the dependence of our models on primary research data. In Section 2, we address the exact meaning of wallet, surveying several different definitions and their implications. We discuss evaluation of wallet models in Section 3 — this is challenging because the wallet is not observed in the larger customer population. In Section 4 we introduce our novel modeling approaches, and demonstrate their results on a data set of IBM customers.

2 Definitions of customer wallet

The first question we need to address is, what exactly is meant by a customer’s wallet? We discuss this in the context of IBM as a seller of IT products to a large collection of customers for whom we wish to estimate the wallet. We have considered three nested definitions:

1. The total spending by this customer in the relevant area. In IBM’s case, this would correspond to the total IT spending by the customer. We denote this wallet definition by TOTAL.
2. The total attainable (or served) opportunity for the customer. In IBM’s case, this would correspond to the total spend by the customer in IT areas covered by IBM’s products and services. While IBM serves all areas of IT spending — software, hardware and services — its products do not necessarily cover all needs of companies in each of these areas. Thus, the served opportunity is smaller than the total IT spending. We denote this definition by SERVED.
3. The “realistically attainable” wallet, as defined by what the “best” customers spend. This may be different from SERVED, because it may just not be practically realistic to get individual customers to buy every single served product from IBM. Defining the best customers is a key in correctly applying this definition. In what follows, we define best customers in a relative sense, as ones who are spending as much as we could hope them to, given a stochastic model of spending. Thus, a good customer is one whose spending is at a high percentile of its “spending distribution”. We describe below some approaches that can allow us to build models that predict such a high percentile and, perhaps more importantly, evaluate models with regard to the goal of predicting percentiles of individual spending distributions. We denote this definition by REALISTIC.

In principle, the three definitions should line up as $\text{REALISTIC} \leq \text{SERVED} \leq \text{TOTAL}$. An important caveat is that all the three definitions could be affected by marketing actions. Thus, a company could theoretically be convinced by diligent marketing activity to buy an IT product they were not planning to spend any money on. This could affect the value of all three wallet definitions. For the rest of this paper, we ignore this possible effect of such marketing actions, and essentially assume that our marketing actions can affect only wallet share, not the actual wallet. Our current challenge is to model these fixed wallet values. We concentrate our interest on REALISTIC and SERVED, which we view as the two more operational definitions. The modeling approaches we discuss below are for the REALISTIC wallet, but we are also developing an approach to modeling SERVED.

We usually know the total company sales (revenue) of potential customers, and the total amount of historical sales made by IBM to these customers (see Section 4 for details). In principle, the relation $\text{IBM SALES} \leq \text{WALLET} \leq \text{COMPANY REVENUE}$ should hold for every company and all three measures (for the REALISTIC definition of wallet, we actually expect $\text{IBM SALES} \approx \text{REALISTIC WALLET}$ for a small percentage of companies).

3 Evaluation approaches

A key issue in any wallet estimation problem is, how can we evaluate the performance of candidate models? Since wallet is usually an unobservable quantity, it is critical to have a clear idea how model performance can be measured — at least approximately — to give us an idea of future performance.

Evaluation using survey values. The first and most obvious approach is to obtain “actual” values for customer wallets, typically through a survey, where customers (or potential customers) are called up and asked about their total IT spend for the previous year. We have at our disposal one such survey, encompassing 2000 IBM customers or potential customers. However, the attempts by us, as well as other groups, to use this survey for model evaluation have not been particularly successful. We attribute this to three main reasons:

1. Of our three wallet definitions, these self-reported wallets correspond to TOTAL, which is the least relevant definition for marketing purposes.
2. Companies have no obligation to carefully consider their answers to such surveys. The exact IT spend may be difficult to calculate, and thus even companies that do not intend to miss-report their wallet may give vastly erroneous numbers.
3. Getting surveys which represent “unbiased” samples of the customer universe is often extremely difficult, due to sampling issues, response bias issues, etc.

The prohibitive cost of surveys compounds these difficulties, and increases our motivation to design methodologies that do not depend on primary research.

Evaluation of high-level indicators. The second common approach is to evaluate wallet models by comparing high-level summaries of model predictions to known numbers, such as industry-level IT spending, or total spending with IBM, and thus evaluating the high-level performance of models. Such approaches include: Comparing the total wallet in an industry to a known “industry opportunity” based on econometric models; Comparing total wallet to the total spending with IBM for large segments of companies and comparing to some “reasonable” numbers; Comparing total wallet in a segment to the total sales of companies in that segment. These evaluation approaches suffer from lack of specificity, as they do not evaluate the performance of the models on individual companies, but rather concentrate on comparing aggregated measures. A more specific approach in the same spirit is based on the statistics of ordering between model predictions and known quantities such as COMPANY REVENUE and IBM SALES. If many of the wallet predictions are smaller than IBM SALES for the same companies or bigger than COMPANY REVENUE, this is evidence of the inadequacy of the model being evaluated.

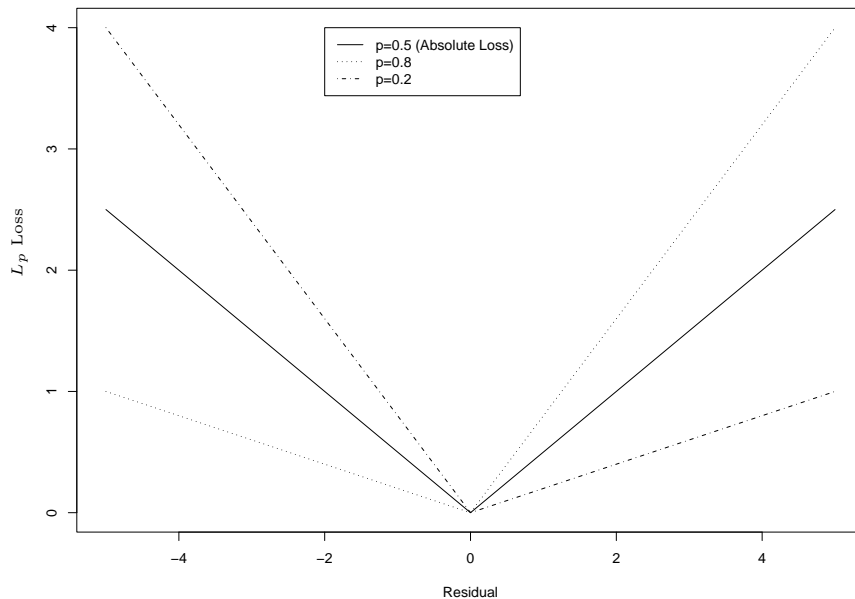


Figure 1: Quantile loss functions for various quantiles.

Quantile loss based evaluation. A different approach is to search for evaluation loss functions which evaluate the performance of the model in a way that is directly

related to the goal of wallet estimation. Despite the fact that the wallet is unobservable, this evaluation turns out to be possible when the REALISTIC definition of wallet is used, using the quantile regression loss function [6]: given an observed IBM SALES number y and a predicted REALISTIC wallet \hat{y} , we define the quantile loss function for the p th quantile to be:

$$L_p(y, \hat{y}) = \begin{cases} p \cdot (y - \hat{y}) & \text{if } y \geq \hat{y} \\ (1 - p) \cdot (\hat{y} - y) & \text{otherwise} \end{cases} \quad (1)$$

In Figure 1, we plot the quantile loss function for $p \in \{0.2, 0.5, 0.8\}$. With $p = 0.5$ this is just absolute error loss. Expected quantile loss is minimized by correctly predicting the (conditional) p th quantile of the residual distribution. That is, if we fix a prediction point x , and define $c_p(x)$ to be the p th quantile of the conditional distribution of y given x :

$$P(y \leq c_p(x)|x) = p, \quad \forall x$$

then the loss function is optimized by correctly predicting $c_p(x)$:

$$\arg \min_c E(L_p(y, c)|x) = c_p(x)$$

With $p = 0.5$, the expected absolute loss is minimized by predicting the median, while when $p = 0.8$ we are in fact evaluating a model’s ability to correctly predict the 80th percentile of the distribution of y given x — exactly the definition of the REALISTIC wallet! Thus, we have a loss function which allows us to directly evaluate model performance in estimating REALISTIC wallets. We will use this loss function $L_p(x)$ in the next section, both for model training and for model evaluation.

4 Modeling approaches

The data we use for modeling is comprised of two main components:

1. **Firmographics.** We derive this data from Dun & Bradstreet¹, which carries information on all businesses in the US and around the world. The data includes:
 - Company size information: revenue and number of employees. Also some information on dynamics of change in these figures in recent years.
 - Various level of industrial classification: industrial sector, industry, sub-industry, etc.
 - Company location: city, state, country.
 - Company structure descriptors: legal status, location in corporate hierarchy (headquarters/subsidiary), etc.

¹<http://www.dnb.com>

2. **Relationship variables.** The IBM relationship variables consist of detailed information about historical purchases made by each IBM customer in the different product categories, as well as other interactions between IBM and its customers.

Our data set contains several tens of thousands of IBM customers (we omit the exact number for confidentiality reasons) who have had business with IBM in 2002 and 2003, together with their firmographics and their IBM Relationship variables for the years 2002-2004, including their total spending with IBM. Our goal for this modeling exercise is to predict their 2004 wallet (although the models will eventually be applied to predict future year wallets, of course). Since we have numerous examples, we leave out 20000 for testing our models, and use the rest as data for training. Throughout our analysis we transform all monetary variables — in particular, the 2004 IBM SALES, used as response — to the log scale, due to the fact that these numbers have very long tailed distributions, which implies that evaluation would be dominated by the few largest companies in the data set. It has often been observed that monetary variables have exponential-type distributions and behave much better when log-transformed (cf. the often cited “pareto rule”).

We discuss here in detail two novel modeling approaches and their performance on this dataset. These approaches aim to model the REALISTIC wallet by utilizing existing machine learning approaches — in this case, nearest neighbor modeling and quantile regression. We are also working on a third approach, which adapts a regression model to estimate either the SERVED or REALISTIC wallet, by relying on some conditional independence assumptions. Details on this modeling approach will be given in future papers.

The REALISTIC wallet definition suggests that some subset of “good” customers already spend close to their total wallet with IBM. Under this assumption, we have a subset of observations for which the wallet is known and could be used as training sample to build a predictive model. It is however impossible to decide *a priori* which subset of firms are those good customers that spent their total wallet with IBM. Consider for instance the naive rule of selecting the top $p\%$ firms with the largest absolute value of IBM SALES. This approach would simply identify the biggest firms. The problem is that the absolute value of the wallet is dependent on firmographics such as size and industry, as well as potentially on the company’s historical relationship with IBM. Without knowing the effect of these variables on the wallet it is impossible to identify these “good” customers.

K-nearest neighbor approach. A reasonable approach towards achieving this relativity effect is through k-nearest neighbor (k-NN) approaches, since the “good” customers would have higher IBM SALES than similar companies. We compare each company to its immediate neighbors in the (historical IBM relationship, firmographics) space, and consider the neighbors with the highest IBM SALES to be “good” customers who are close to spending their whole IT wallet with IBM. The

main firmographic variables we use in our k -NN implementation are industry and company size (as indicated by the D&B employee and/or revenue numbers). We also use the previous year (in this case, 2003) total IBM SALES. The wallet of a firm A is estimated by first ranking all other firms based on their similarity to A , selecting a neighborhood of k firms that are very similar to A and finally ranking the k firms by their IBM SALES. Following the initial argument that the best customers spent all their attainable wallet with IBM, we can predict a wallet as a statistic on the ranked IBM SALES of these k firms. A case can be made for taking the MAXIMUM as the appropriate statistic, however this measure is very volatile and tends to result in overly optimistic prediction. A more practical approach is to take the REALISTIC definition of wallet and apply it directly, by taking the p th quantile of these neighbors as our estimate for firm A .

We have implemented this strategy and evaluated different neighborhood sizes and different variables for the similarity definition. By far the best neighborhood definition we found was using the $\log(2003 \text{ IBM SALES})$ for the distance while grouping on the industry. In other words, A 's neighbors are the companies in the same industry with the closest IBM SALES in 2003.

In our experiments we used the $p = 0.9$ quantile of the 2004 IBM SALES in a company's neighborhood in the training data as the prediction of its 2004 wallet. We experimented with various neighborhood sizes, ranging from 10 to 400. Our evaluation on the 20000 test samples uses the quantile loss $L_{0.9}$ of equation (1) on the log scale. Figure 2 shows the mean loss incurred by these models on the 20000 test set observations (the solid line marked k -NN). A "naive" baseline model to compare these results to is simply predicting the 2003 IBM SALES as the 2004 wallet, which gives a mean test error using $L_{0.9}$ of 0.365. Figure 2 also shows the *quantile regression* results discussed below. The model using 50 neighbors is clearly the best and it achieves about 25% lower error on the holdout data than the naive model.

It is also interesting to investigate the residuals of these models. Figure 3 (left panel) shows a sample of 1000 test-set predictions of the model with $k=50$ neighbors, plotted against the 2004 IBM SALES. We see that most of the predictions are higher than the sales, as we would expect from a wallet prediction model. In fact, the use of $p = 0.9$ in the quantile loss implies we would expect about 90% of predictions to be higher than the actual number. Two sets of observations stand out in this scatter plot:

- The ones with no 2004 IBM SALES which form the vertical column of points on the extreme left. Despite these customers buying nothing from IBM in 2004, our model still predicts a non-zero wallet for all of them, for some even a substantial wallet (as high as \$163K, after exponentiation). This is completely consistent with the concept of REALISTIC wallet, which models what the customer *could* spend, as opposed to actual spending.
- Four companies get a prediction of 0. These companies have no 2003 IBM

SALES, and consequently so do many of their neighbors in the training data. In these four neighborhoods, it happens that the 90% 2004 IBM SALES among the neighbors was also 0, and thus so was the prediction. It should be noted that these four companies *are not* the only ones with no 2003 IBM SALES, and the other companies did get non-zero predictions, because their neighborhoods contained companies that did buy from IBM in 2004.

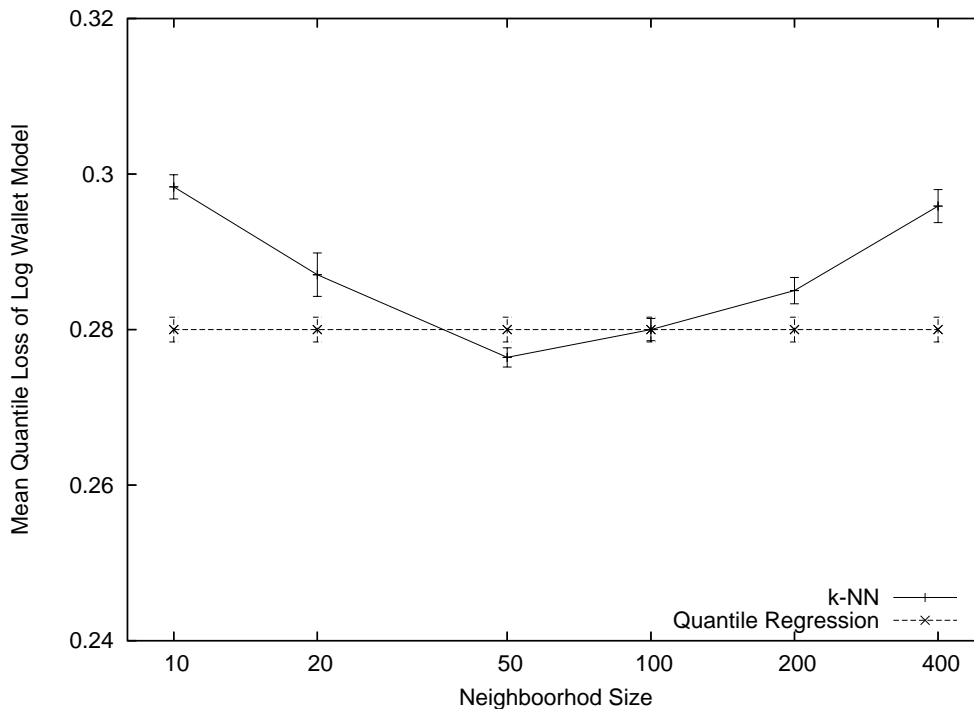


Figure 2: Quantile loss of k-NN and quantile regression models in predicting 2004 wallet. Also shown are 95% confidence bars on the performance.

The right panel of Figure 3 shows the histogram of the residuals, and we observe that about 86% of the test set observations (17183/20000) generate predictions that are higher than actual 2004 IBM SALES. As discussed in Section 3, we can consider this to be a “high level summary” which indicates the reasonable performance of our approach.

Quantile regression approach. Another way of accounting for the relative definition of “good” customers is to treat this as a predictive modeling problem, but model the p th quantile of the conditional spending distribution of y (future IBM SALES) given x (firmographics and relationship variables). As we have seen in Section 3, the quantile regression loss function L_p of equation (1) is optimized by predicting this p th quantile. This loss function can be used for training as well. This is the

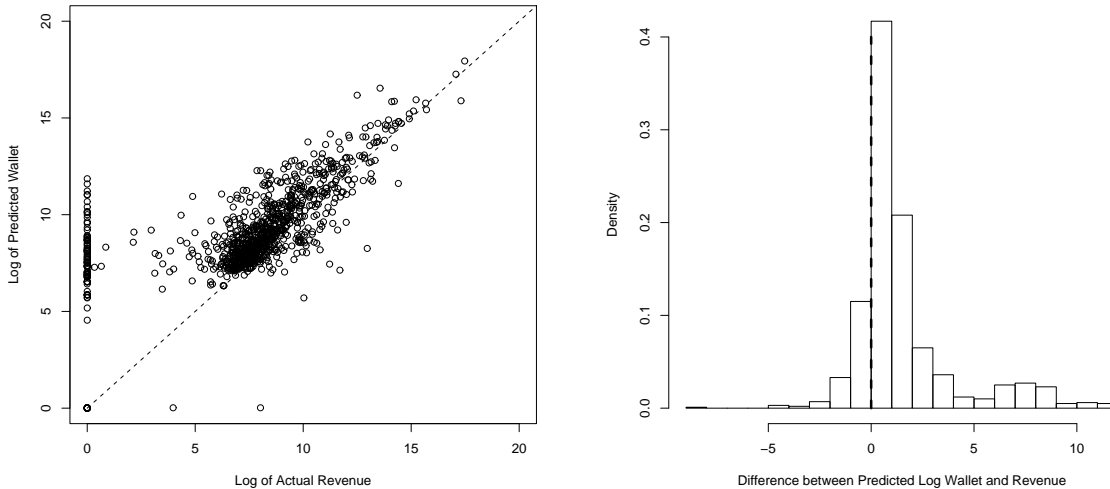


Figure 3: A sample of model predictions from the 50-NN model, plotted against 2004 IBM SALES (left) and a histogram of test set residuals obtained from this model, i.e., $model\ prediction - 2004\ IBM\ SALES$ (right), demonstrating that about 90% of predictions are higher than actual sales, as we would expect.

method typically known as *Quantile Regression* [5, 6]. Given a set of training data $\{x_i, y_i\}_{i=1}^n$, linear quantile regression seeks the linear model in \mathbf{x} which minimizes the empirical quantile error L_p :

$$\hat{\beta} = \arg \min_{\beta} \sum_i L_p(y_i, \beta' x_i)$$

The R package `quantreg`² implements linear quantile regression and is the one we used for our experiments below. It should be noted that due to issues such as over-fitting and bias, it is not guaranteed that using the “correct” loss function for modeling will result in a good prediction model for future data with regard to the same loss function. It can be shown, however, that this loss function is “consistent”, i.e., with infinite training data and correct model specification it will indeed result in the p th quantile of $P(y|x)$ as the model predictions.

We have experimented with several quantile regression models, varying the choice of independent variable x chosen for the model. With our large training data set, we concluded that the saturated main effect model, with no interactions, is the best one to use. This model included numerous firmographic and IBM relationship variables, with a total of 45 degrees of freedom. The test set performance of this model, using the $L_{0.9}$ loss function, is shown in Figure 2. We see that it is comparable to the

²For details on R, see www.R-project.org

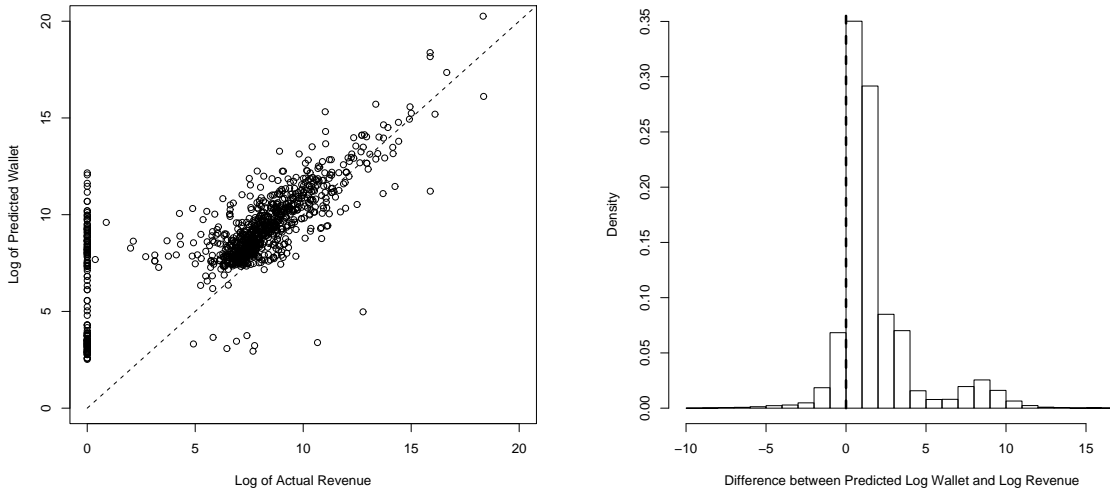


Figure 4: Same plots as in Figure 3, for the quantile regression model. The conclusions are similar.

best k-NN model, using $k=50$, but slightly inferior (0.280 vs. 0.276). Given the very large size of our test set, this difference is marginally significant (t-test p-value 0.02), but it is clearly too small to be considered meaningful.

Figure 4 shows the residual analysis of the test set performance of the quantile regression model, in the same spirit as Figure 3. The conclusions are also similar. Note that there are no 0 predictions from this model, but the cloud of low predictions now contains nine points, as compared to four for the nearest neighbor model. This is due to the fact that the 2003 IBM SALES play a dominant role in determining model prediction. Thus, all companies with no 2003 IBM SALES get very low predictions, although none are exactly zero. Overall, the percentage of test set data points for which the prediction is above the 2004 IBM SALES is almost exactly 90% (18009/20000). This supports our conclusion that our model is doing what we expect and suffers little or no overfitting.

5 Summary

We have proposed three definitions for customer wallet: TOTAL, SERVED and REALISTIC, and argued that the last two are the ones that should be modeled. We have described some of the difficulties in evaluating performance of wallet models without relying on primary research results, which are expensive, of questionable quality and usually are relevant only for TOTAL wallets. In this context, we have

suggested the use of the *quantile regression* loss function for evaluating the REALISTIC wallet predictions.

Our main focus is on developing predictive modeling approaches for SERVED and REALISTIC wallet estimation. We have proposed and discussed in detail two new methodologies for modeling the REALISTIC wallet: quantile nearest neighbor and quantile regression. Our experiments in applying these approaches to actual data show promising results. Our future work will report on our third methodology, the “decomposition” approach, which adapts a standard regression model to estimate the either the SERVED or REALISTIC wallet, by relying on some conditional independence assumptions.

Acknowledgements We thank Alexey Ershov and John Waldes from IBM for useful discussions.

References

- [1] Du, Rex, Thesis proposal, Fuqua School of Business, Duke University, 2005.
- [2] Epsilon Data Management, Solving the Customer Value Puzzle, 2001.
<http://www.dmreview.com/whitepaper/WID384.pdf>.
- [3] Garland R., Share of wallet’s role in customer profitability, *Journal of Financial Services Marketing*, volume 8, number 8, pp. 259-268(10), 2004.
- [4] Keiningham T.L., Perkins-Munn T. and Evans H., The Impact of Customer Satisfaction on Share-of-Wallet in a Business-to-Business Environment, *Journal of Service Research*, volume 6, number 1, pp. 37-50(14), 2003.
- [5] Koenker, R. *Quantile Regression*. Econometric Society Monograph Series, Cambridge University Press, 2005.
- [6] Koenker, R. and Hallok, K. Quantile Regression: An Introduction, *Journal of Economic Perspectives*, 2001.
- [7] Malthouse, E. C. and Wang, P. Database Segmentation using Share of Customer, *Journal of Database Marketing*, volume 6, number 3, pp. 239-252, 1998.
- [8] Pompa N., Berry J., Reid J. and Webber R., Adopting share of wallet as a basis for communications and customer relationship management, *Interactive Marketing*, volume 2, number 1, pp. 29-40(12), 2000.
- [9] Rosset S., Neumann E., Eick U. and Vatnik N., Lifetime Value Models for Decision Support. *Data Mining and Knowledge Discovery Journal*, Vol. 7, 321-339, 2003.
- [10] Rosset, S., Perlich, C. and Zadrozny, B. Ranking-Based Evaluation of Regression Models. *IEEE Conference on Data Mining 2005*, to appear.