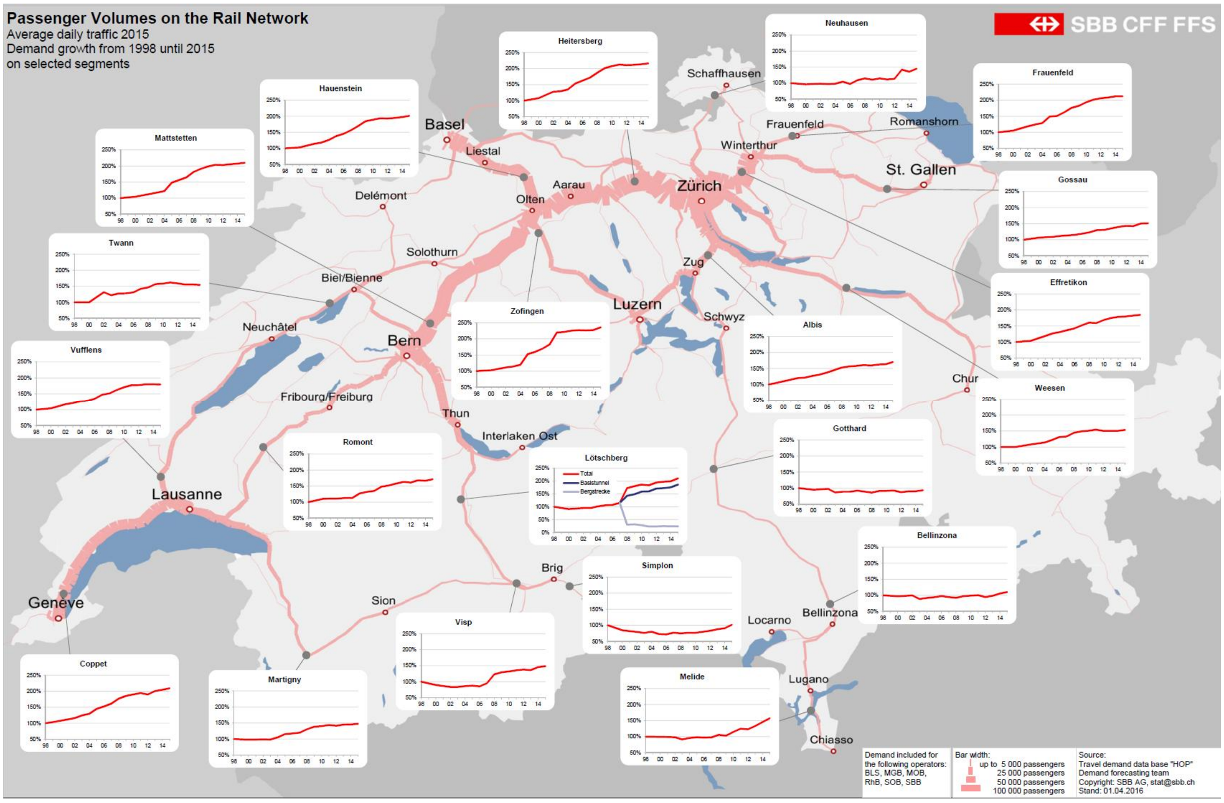


Calibration of the Swiss Rail Model Using Longitudinal Demand Observation

Wolfgang Scherr, SBB
Patrick Bützberger, SBB

Conference paper STRC 2016

STRC | **16th Swiss Transport Research Conference**
Monte Verità / Ascona, May 18-20, 2016

Calibration of the Swiss Rail Model Using Longitudinal Demand Observation

Wolfgang Scherr
SBB
Passenger Division, Corporate Development
Bern, Switzerland

Phone: +41 79 944 95 63
Fax: +41 51 220 46 98
email: wolfgang.scherr@sbb.ch

Patrick Bützberger
SBB
Passenger Division, Corporate Development
Bern, Switzerland

Phone: +41 79 753 41 65
Fax: +41 51 220 46 98
email: patrick.buetzberger@sbb.ch

May 2016

Abstract

This paper describes the calibration process of the travel demand model that is used at Swiss Federal Railways (SBB). The model covers intercity and regional demand. It is a direct demand model which draws service quality indicators from a dynamic passenger assignment.

For the recent recalibration of the demand model parameters, a longitudinal dataset with 160'000 origin-destination pairs was developed. It contains consistent data for three years (2004, 2007, 2012), covering travel demand, transport supply for rail and road, and socio-economic variables (population, GDP). Model parameters were estimated using a subset of 12'000 origin-destination pairs. Then the entire longitudinal dataset was used to measure prediction accuracy; i.e. how close the observed change in travel demand is to the simulated change in demand. Prediction accuracy served as a major criterion to select the optimal model.

As a result of this study, the effectiveness of forecasts has been improved. The results confirm that demand elasticities are variable and depend on origin-destination characteristics such as trip length and level of service. The paper concludes with an outlook to further model development covering aspects such as demographic change, competition on transportation markets, and emerging transport technologies.

Keywords

travel demand model – calibration – longitudinal observation – elasticity – rail – passenger demand – ridership forecast – prediction accuracy – prediction error – validation

1. Introduction

In practice, most travel demand models are calibrated using cross-sectional data. The cross section is taken for the base year, and calibration is limited to comparing the observed reality of the base year to the simulation results. In this paper we present a calibration process which uses longitudinal observations. Longitudinal data allow to measure prediction accuracy and thus to test the ability of a model to make a good forecast. The time period covered by this longitudinal dataset is eight years long, from 2004 to 2012. During this period major changes in rail travel times and level of service happened all across Switzerland. Hence, this data basis allows to observe how travel demand has reacted to changes in rail service quality. As a result of the work with the longitudinal data set, new model parameters were estimated, prediction accuracy was improved, and assumptions on level-of-service elasticities of rail passenger demand were revised.

Travel forecasts are criticized for their inaccuracy. Comparing predicted versus observed demand in over 200 international projects between 1969 and 1998, Flyvberg, Næss et al. (2005, 2006) find, that one third of the projects show a forecasting error of more than +/- 40% and that inaccuracy is worse for rail projects than for road projects. More recently, Hartgen (2013) analyzes current U.S. and international practice, finding that most 20-year travel demand forecasts show errors of at least +/-30%, and concludes that technical and institutional improvements are necessary to improve accuracy and therefore the usefulness of travel forecasting.

Forecast accuracy[1] stands for the question: how well does a model predict the future? In econometrics, this is also called the external validity of a model (how good does the model predict scenarios that are different from the observed state) as opposed to the internal validity (how well does the model explain and reproduce the observed reality). Unfortunately, the practice of travel model validation is almost completely limited to the internal validity, based on cross-sectional data: Sammer et al. (2014) document this common practice for the German speaking countries. If the prediction accuracy of travel models is tested, then it is typically limited to ex post analysis of individual projects (Hartgen 2013).

Within Swiss Federal Railways (SBB), ridership forecasts have always been under scrutiny, as management bases business decisions on ridership predictions and there is an expectation

[1] In Flyberg/Naess (2005, 2006), inaccuracy of a travel forecast is defined as actual minus forecasted traffic in percent of the forecasted traffic. In a forecast of public transportation, “traffic” is to be replaced by “passengers”.

that the predictions “have to fit”, especially in the mid-term horizon. In this context, the modeling approach at SBB puts an emphasis on empirical demand data and prediction accuracy is often reviewed for particular rail projects. In the work presented in this paper we go beyond prediction accuracy on the project level, but evaluate prediction accuracy on the system level, and over a long period of time.

This paper has the following outline: Section 2 introduces the model environment at SBB including the concept of the direct demand model and the explanatory variables. The development of the longitudinal database is described in section 3, as well as its use in parameter estimation and verification of prediction accuracy. Section 4 compares model predictions and empirical observations. The results are discussed in section 5 with a focus on how the optimal model was selected. Finally, section 6 draws conclusions and gives an outlook into challenges and approaches of the ongoing model development at SBB.

2. SBB’s ridership forecasting model

2.1 Purpose and main characteristics of SBB’s rail model

The purpose of transportation modeling at the SBB passenger division is to support management decisions about future service concepts and investments in infrastructure and rolling stock. To fulfill this mission, the SBB rail model SIMBA (Olesen et al. 2016) has to explain not only the demand for rail travel but also the production side. On the demand side, the model predicts how ridership will react to changes in rail service, and how these changes will affect revenue. On the production side, the model predicts the need of rolling stock and then the cost of production of a particular service concept. Over the course of a year up to 100 model applications are conducted, for projects in regional, domestic intercity and international rail service. The forecast horizons are mid-term (from 1 year to 6 years) and long-term (up to 25 years). The model is developed in-house at SBB, by the modeling staff of SBB’s corporate development unit in the passenger division.

SIMBA is a macroscopic model on 2’100 zones. It covers Switzerland and all rail corridors into the neighbouring countries. SIMBA uses the software Visum by PTV for time-table development, route-choice and assignment. The prediction and evaluation system is SBB’s own development. Rail supply is modelled with 1’800 rail stations, 650 routes (a.k.a. “time profiles”) and a timetable of roughly 12’000 train trips. The timetable in the model covers 24 hours, is differentiated for weekday and weekend, and consistently coded for existing and future states. The passenger flow model is a 24-hour dynamic assignment, using the “timetable based assignment” method in Visum (Friedrich et al. 2001). The dynamic assignment was calibrated by the SBB modelling team, assignment parameters were optimized to best fit the observed route choice, capacity-constraints are included in route choice (Lieberherr et al. 2012) and a method for time-of-day demand distributions has been developed (Kaeslin et al. 2014). The indicators of service quality per origin-destination pair, such as travel time, are derived from the 24-H dynamic route choice, which allows to predict the impact of time-table changes at specific times of the day. Forecasting of the domestic demand uses a direct demand model (see section 2.2 on page 4 of this paper), while international intercity demand uses a multimodal approach. The demand models are based on survey data with almost 160'000 OD pairs in the empirical demand data base.

2.2 Direct demand model

There are different approaches to ridership forecasting. Most common are macroscopic multi-stage models. An emerging approach uses microscopic models (a.k.a. agent-based models). To forecast domestic passenger demand in SIMBA, a third type is being used, called direct demand model. The term "direct"[2] was chosen to distinguish this class of models from multi-stage models, as direct demand models do not explicitly model destination choice and mode choice. The class of direct demand models can be formulated as follows:

$$T_{od} = a \cdot \prod \left(X_{od}^{i}\right)^{\theta_i}$$

Where T_{od} is the demand of passenger trips (P-Trips) from origin *o* to destination *d*, and *X* are the explanatory variables (with index $i = 1 \ldots k$); θ and a are the parameters of the model. It can be shown that the parameters θ equal the direct demand elasticity $\theta_i = \frac{\partial T/T}{\partial X_i/X_i}$. Hence this kind of incremental model is also called „elasticity model" and the θ_i are referred to as elasticity parameters.

An incremental form of the model is obtained by dividing the same formula of a future state $T_{od}(1)$ by the form of the reference state (or base year) $T_{od}(0)$:

$$T_{od}(1) = T_{od}(0) \cdot \prod \left(\frac{X_{od}^{i}(1)}{X_{od}^{i}(0)}\right)^{\theta_i}$$

The incremental form above is used in SBB's model „SIMBA". It is also prevailing practice in British ridership forecasting (ATOC 2013). The incremental form is popular as it allows for integration of observed OD demand matrices in the forecast.

---

[2] Direct demand models are in use for intercity demand models since the 1960s. An examples of an early direct demand model in intercity transport is Quant-Baumol (1966).

A constant growth factor (*UTG*) can be added as follows:

$$T_{od}(1) = T_{od}(0) \cdot \prod \left(\frac{X_{od}^{i}(1)}{X_{od}^{i}(0)} \right)^{\theta_i} \cdot UTG$$

UTG represents the trend of unexplained growth in the observed time period; i.e. the demand growth that cannot be explained by the variables *X*.

To estimate the parameters θ with linear regression, the incremental model can be linearized as follows:

$$\ln \left(\frac{T_{od}^{i}(1)}{T_{od}^{i}(0)} \right) = \sum \left[\theta_i \cdot \ln \left(\frac{X_{od}^{i}(1)}{X_{od}^{i}(0)} \right) \right] + intercept$$

Now, the unexplained trend growth can be derived from the constant (*"intercept"*) and from dummy variables as follows: *UTG = exp(intercept + dummy)*.

2.3 Explanatory variables

Table 1 shows the explanatory variables currently used in SIMBA's direct demand model, together with the elasticity parameter θ which had been used previously for each variable.

Table 1 Explanatory variables in SIMBA's direct demand model

X_i	variable content	type	θ_i
TT	travel time, rail	endogenous	*-1.0*
NT	number of transfers, rail	endogenous	*-0.1*
AT	departure adaptation time, rail	endogenous	*-0.4*
TA	tarif, rail	endogenous	*-0.4*
Pop	population + employment	exogenous	*+1.0*
GDP	gross domestic product (per capita, in real terms)	exogenous	*+0.4*
TT_RD	travel time, road	exogenous	*+0.6*
UG	unexplained growth trend	exogenous	*+1.0*

All the models in this paper are based on the variables above. During this project, additional variables were added but had to be discarded because of input data insufficiencies or because parameter estimation did not produce significant results.

2.3.1 Endogenous variables: service quality indicators

The three indicators of rail service quality are defined as follows:

- Travel time (*TT*) is defined as time from departure at the origin zone to arrival at the destination zone – it includes in-vehicle time and out-of-vehicle time, such as transfer waiting time and walk times, but not the waiting time at departure.
- Number of transfers (*NT*) has the purpose to measure the directness of the service.
- Departure adaptation time (*AT*) is the difference between desired departure time and actual departure time. It serves to measure service frequency. It is computed as average over all passengers over 24 hours and hence reacts to modifications of service headways at any time of day.

These three indicators are computed based on the route choice of SIMBA's 24-hour dynamic assignment, as average over all connections r, and averaged over weekday and weekend. As an example, the formula for travel time TT is given:

$$TT(o,d) = \sum_{r\ from\ o\ to\ d} TT(r) \cdot p\_trips(r) \Big/ \sum_{r\ from\ o\ to\ d} p\_trips(r)$$

Only recently, the computation of service quality indicators based on a 24-hour dynamic assignment has been put into practice at SBB. When this project of model calibration took place, we also analyzed a previously applied method where the service quality indicators are computed based on static route choice. In this paper we only show results of estimation and prediction success for the new method, i.e. indicators from the 24-hour dynamic assignment.

The forth endogenous variable is tarif TA: It stands for a nationwide rail travel price index, which is the average over all ticket types including monthly or annual passes.

2.3.2 Exogenous variables

All socio-economic variables are derived from national statistics provided by from either the Federal Statistics Bureau (BFS) or the Federal Land Use Agency (ARE). The granularity of the raw data is different for each variable. For population and employment, data come at high granularity and are aggregated to SIMBA's traffic zones. Other variables, like GDP, are available only on the level of cantons; in these cases, the direct demand model permits to use the index of variation *X(1)/X(0)* on a higher geographic level. Road travel times are computed on an origin-destination level using the national travel model of the Federal Department UVEK.

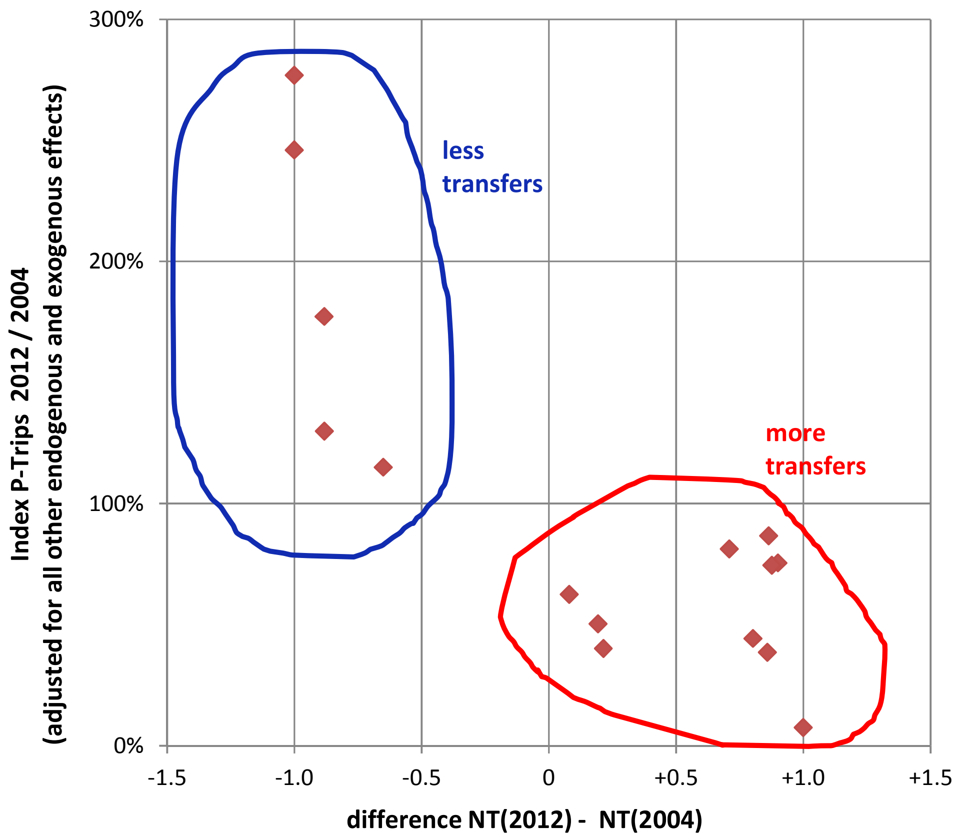
3. Methods

This section of the report explains how we measured demand and explanatory variables for the time period from 2004 to 2012, by building the longitudinal database, then using a sub-sample of the database for parameter estimation, and finally by testing the different models and evaluating their prediction success.

Rail demand in Switzerland has consistently grown in the past: PKM have almost doubled over the last 20 years. In the time period from 2004 to 2012, which we deal with in this paper, P-Trips have grown by around 50%. During the same time, rail service has been improved significantly, and exogenous drivers (population and economy) have grown strong as well. Hence, the time period is perfectly suited to observe how the different endogenous and exogenous variables influence rail ridership.

Figure 1 shows for selected OD pairs, how changes in the service quality indicator “directness” are related to changes of passenger demand (adjusted for other effects).

Figure 1 Ridership growth versus changes in service quality for selected OD pairs



3.1 Longitudinal database

The longitudinal data set covers three cross sections, namely the years 2004, 2007 and 2012. The data in each year describe the Swiss rail system as a whole, with demand and supply, and the exogenous environment. This data set allows to put the variations of travel demand in contrast to variations of all explanatory variables. We looked at two prediction periods: 2004 → 2012 and 2007 → 2012 [3].

The challenge in building the data base is to obtain the information from the same sources, with the same data granularity, and to use identical data processing methods for all three years. In cases where methods have improved over the years, we have either applied the newest method backward to the previous years; or if that was not possible, we recalculated the data of the later years with the outdated method.

The following data are included:

- Person trips (P-Trips):
  Average day travel demand on 155'000 OD pairs. This empirical trip table is derived from the annual on-board survey which produces around 7 million OD records per year. The demand is expanded with passenger count data to obtain the trip table.
- Rail timetable and rail service:
  A Visum network model with rail timetables, built with the same methodology for all three years, and hence allowing to compute route choice and service quality indicators for each OD in a consistent manner.
- Rail tarif level.
- All exogenous variables (see section 2.3.2).

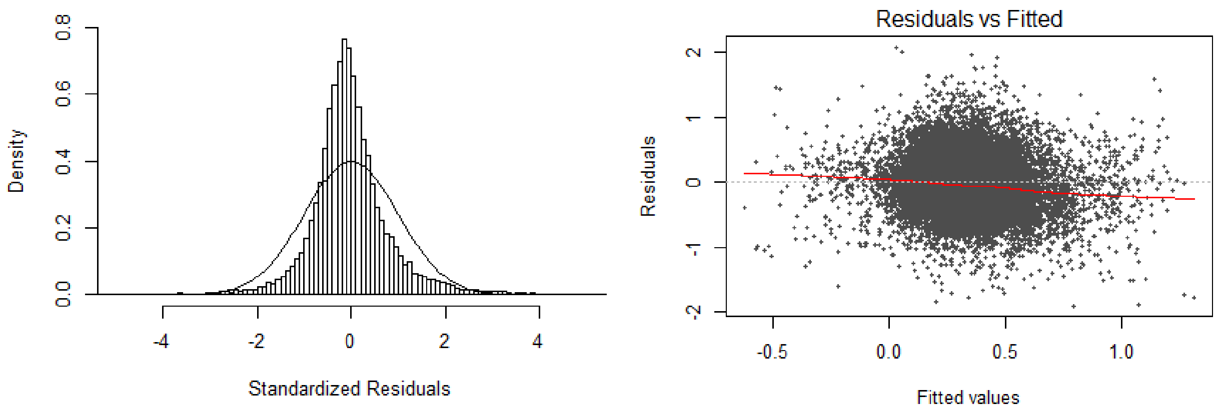
3.2 Model estimation

Estimation of elasticity parameters was not performed on the entire database, but on a sub-sample of approximately 12'000 OD pairs, which had a statistically sufficient sample size of interviewed passengers. Each OD pair functioned then as one observation point in the estimation.

[3] In theory, we could have analysed 2004 → 2007. However, that period would not be useful, because a major service improvement program called "Bahn 2000" (including the opening of the fast track Bern-Olten) took place in 2005, which triggered a big demand reaction that had not yet been completed in 2007.

The classical ordinary least squares method (OLS) was used, weighted by P-Trips per OD pair. OLS is based on certain assumptions, most importantly a near linear correlation and normally distributed, independent residuals with expected value 0 and constant variance. These assumptions are not perfectly satisfied in the data set (see Figure 2). However, the residual-vs-fitted plot shows that residuals are clustered almost symmetrically around the mean (0), and no clear pattern that would contradict requirements of OLS. In order to further validate the results, other, more robust linear regression models were applied, such as the median regression, also known as least absolute deviation (LAD). These robust methods have the advantage of being less sensitive to violating OLS model assumptions, but they are less efficient. These estimations with the LAD model did not produce significantly different results and thus, the OLS method is considered adequate (see Bützberger 2013).

Figure 2 Residual distribution of the linear regression with OLS



The logarithm of the change in P-Trips is the depending variable in all estimations. For most estimations, P-Trips served also as the weight of each observation. After having completed prediction success testing, some models were re-estimated using PKM as the weight, in an attempt to better represent longer trips in the parameters.

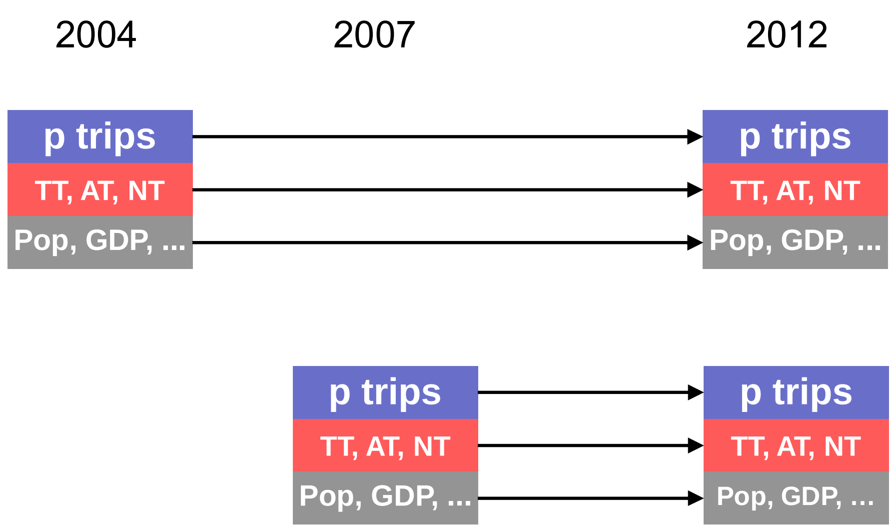
The two observation periods 2004 → 2012 and 2007 → 2012 were first estimated separately. Here we found that rail service elasticities were rather independent of the observation period. Exogenous variables however turned out differently for each period, mainly because the data quality was not consistent over all three years. Therefore, in a second wave of estimations, both periods were combined in one data set and estimations were repeated with fixed parameters for the exogenous variables.

3.3 Test of prediction success

There is no guarantee that parameters derived by statistical estimation will be effective in predicting the future. Our estimations had two limitations: first, we could use only a sub-sample of the OD pairs because of survey sample size. Second, most estimations were performed so that the error in P-Trips was minimized; but for a railway company, prediction success in PKM is more important than in P-Trips.

The concept of our prediction success tests is to make the models predict "ex post" the situation of 2012. Using our longitudinal data base, we had two test periods, a long-term period with 2004 as the base year and a mid-term period with 2007 as the base year.

Figure 3 Concept of model calibration and validation with the longitudinal data set



Several models (i.e. elasticity parameter sets) underwent the test of their prediction success. Each model was tested on three sub samples:

- The estimation sub-sample (12'000 OD pairs)
- An extended sample with all OD pairs that had been consistently surveyed in all three years (140'000 OD pairs)
- The entire passenger demand on all OD pairs (155'000 OD pairs)

All error statistics have been evaluated for both P-Trips and PKM, and for two prediction periods (2004 → 2012 and 2007 → 2012). We used the following error statistics;

- error $\Delta Y := Y_{model} - Y_{observed}$
- inaccuracy $ia := \Delta Y \;/\; Y_{observed}$
- average errors $RMSE := \sqrt{\frac{\sum(\Delta Y)^2}{N}}$ and $MAE := \frac{\sum|\Delta Y|}{N}$

4. Results

This section presents major results that were obtained by applying the techniques outlined in section 3. These results are the raw data and major observations, conclusions will not yet be drawn in this section. There are four sub-sections, one about estimation results and the other three about prediction success.

While we have estimated surely more than 200 models in this project and tested more than 30, we show here only a small selection of five models. These five models represent the situation at the end of our project and illustrate our final choice to pick the "best model" to be used in SIMBA's direct demand model. The following five models are presented:

- m0 – constant elasticity parameters previously used in SIMBA (see Table 1)
- m1 – constant elasticities, derived by estimation
- m2 – constant elasticities, stratified by OD clusters (e.g. intercity, agglomeration, …)
- m3 – linear-variable elasticities, estimated with observations weighted by P-Trips
- m4 – linear-variable elasticities, estimated with observations weighted by PKM

4.1 Estimated parameters

Table 2 shows the estimation results for the four models m1, m2, m3 and m4. For these estimations, the two periods 2004 → 2012 and 2007 → 2012 were combined in one observation data set. The estimations of these four models took place at the end of this project. At that point in the project, we fixed (pre-defined) the parameters for the exogenous variables so that the regression procedure could not determine their values. We had obtained the fixed parameters with earlier estimations on a partial set of OD pairs with sufficient quality. The explanatory variable rail tarif (*TA*) does not appear because at the time of the project, we had no data with sufficient granularity to support parameter estimation.

The important difference between these models lies in the parameters for rail service quality (TT, AT, NT) and how they vary for different situations. In model m0 and m1 they are constant for the entire system. In m2 they vary for different types of OD pairs (e.g. intercity, inside of agglomerations, to/from agglomerations). For m3 and m4, the variation is computed as a linear function of travel time (TT) and service frequency (AT). We also estimated models where the variation is quadratic, but they were not retained (see section 5.1 on page 18).

Table 2 Linear regression results for four models

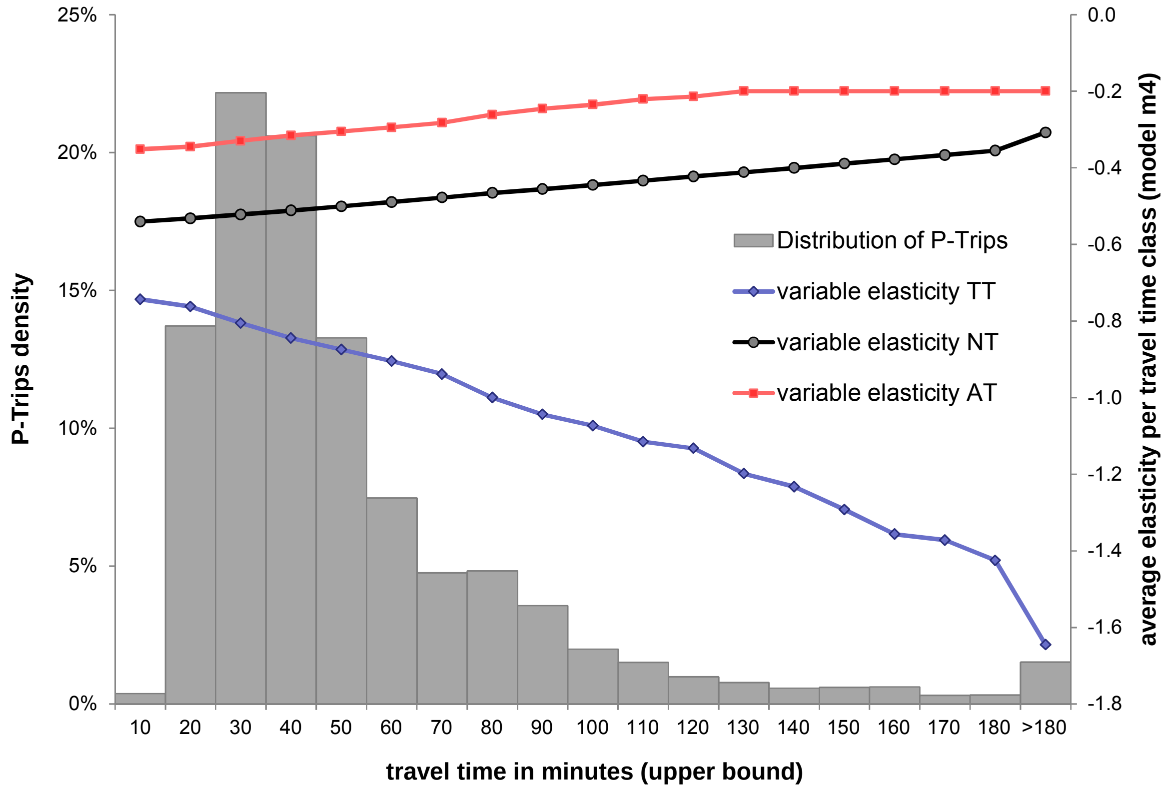
	m1	m2	m3	m4
Endogenous variables:				
TT_rail	**-0.830**	**-0.579**	**-0.702**	**-0.813**
TT_AT/15			**+0.300**	**+0.300**
TT_TT/45)			**-0.269**	**-0.232**
TT(Intercity)		**-0.549**		
TT(fromto_Agglo)		**-0.602**		
NT_rail	**-0.481**	**-0.450**	**-0.557**	**-0.551**
NT_TT/45			**+0.067**	**+0.052**
AT_rail	**-0.230**	**-0.238**	**-0.352**	**-0.330**
AT_AT/15			**-0.150**	**-0.100**
AT_TT/45			**+0.231**	**+0.080**
AT(Agglo)		**-0.094**		
AT(Intercity)		**+0.189**		
AT(fromto_Agglo)		**+0.134**		
Exogenous variables:				
GDP	***+0.500***	***+0.500***	***+0.500***	***+0.500***
Pop	***+1.700***	***+1.700***	***+1.700***	***+1.700***
TT_road	***+0.600***	***+0.600***	***+0.600***	***+0.600***
Dummies for OD type / intercept:				
AggloBS (intercept)	**+0.099**	**+0.092**	**+0.096**	**+0.067**
AggloZH	**+0.086**	**+0.086**	**+0.087**	+0.010
AggloLeman	**+0.127**	**+0.122**	**+0.124**	**+0.068**
AggloTI	**+0.232**	**+0.243**	**+0.234**	**+0.199**
R_D-CH	**+0.059**	**+0.066**	**+0.055**	**+0.047**
R_Romand	**+0.041**	**+0.051**	+0.034	**+0.094**
fromto_Agglo	**-0.027**	**-0.018**	**-0.021**	-0.009
InterCity	+0.021	**+0.046**	**+0.042**	**+0.029**
Röstigraben	-0.004	+0.009	-0.005	**-0.040**
Periphery	**-0.024**	-0.016	**-0.026**	**-0.043**
Airport	**+0.210**	**+0.218**	**+0.221**	**+0.145**
Fit statistics				
R2	0.203	0.211	0.218	0.340
adjR2	0.202	0.210	0.218	0.339
MSE	8.90	8.81	8.75	222.18
RMSE	2.983	2.968	2.958	14.906
SSE	196757	194698	193433	4911104
DF	14	19	17	15
N	22119	22119	22119	22119
Weight	P-Trips	P-Trips	P-Trips	PKM

Notes:
Significance levels: bold : $p < 0.01$, normal: $p > 0.01$
Dependent variable: ln([P-Trips(1)/ P-Trips(0)]), The logarithm is also applied to all explanatory variables.
Parameters for the exogenous variables are fixed.

Note that the depending variable of all estimated models in table Table 2 is the logarithm of the change in P-Trips, or ln([P-Trips(1) / P-Trips (0)]). The estimations differ in their weight, which is either P-Trips or PKM. Hence the statistics MSE, RMSE and SSE, which are computed weight-dependent by the software R (2008) cannot be compared between m4 and the other models.

Figure 4 shows how P-Trips are distributed by classes of travel time, and how the elasticities for model m4 vary as a function of travel time.

Figure 4 Distribution of P-Trips and variable elasticities in travel time classes

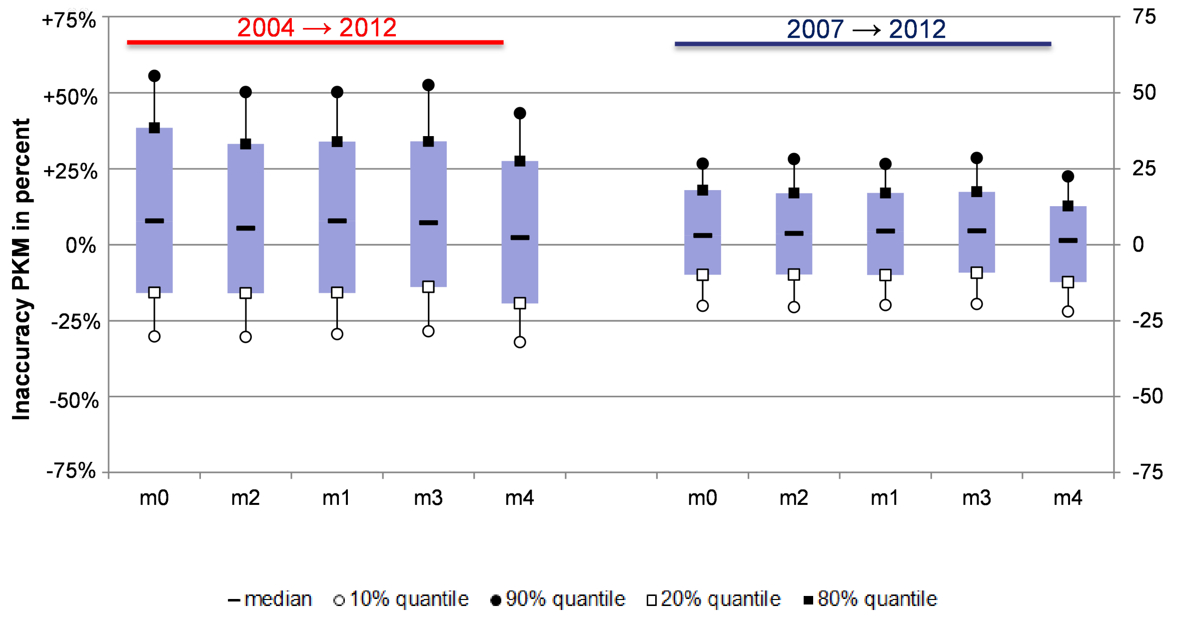


4.2 Prediction success on the estimation subsample

A first series of prediction success tests were performed on the same sample that was used for parameter estimation (roughly 12'000 OD pairs).

Looking at the Box-Plots in Figure 5, which shows test results for both prediction periods 2004 → 2012 and 2007 → 2012, it can be seen how the five models vary in the range and distribution of inaccuracy. Model m4 turns out to have the most balanced range in predicting PKM, and the median clos→→est to 0. Further evaluations showed that m4 did also perform equal or better in inaccuracy box-plots for P-Trips, which are not shown in this paper.

Figure 5 Prediction inaccuracy (PKM), 5 models, estimation sub-sample (12'000 OD)



The following table shows an aggregate statistic for prediction success, the average error (measured as RMSE[4]) over all OD pairs, both for PKM and P-Trips. All estimated models perform better than the "old" model m0, but m4 stands out with the smallest errors.

Table 3 Prediction Success – RMSE – estimation sub-sample (12'000 OD pairs)

		m0	m1	m2	m3	m4
2004 → 2012	P-Trips	82.4	75.5	77.9	75.8	68.6
2007 → 2012	P-Trips	46.3	44.9	45.2	45.2	39.8
2004 → 2012	PKM	3175	2811	2957	2928	2459
2007 → 2012	PKM	1493	1496	1666	1569	1282

[4] Note that the RSME values in Table 2 are not comparable with those in Table 3. The former uses ln[P-Trips(1)/P-Trips(0)] as dependent variable, while the latter uses P-Trips as dependent variable. Another difference is that the former computes weighted square error, while the latter uses the more common un-weighted form (as defined in section 3.3 on page 10).

4.3 Prediction success on the extended sub-sample

The second set of prediction success tests was performed on an extended sample covering roughly 140’000 OD pairs, which have all been consistently surveyed in the time period 2004 through 2012. This sample carries the “problem of small numbers”; i.e. small numbers of passenger trips with a low likelihood of a representative survey result.

The inaccuracy results are similar to the previous section with 12’000 OD pairs: all estimated models outperform the old model, and model m4 performs best in in the distribution of inaccuracy (Figure 6) and in the average error for both P-Trips and PKM (Table 4).

Figure 6 Prediction inaccuracy (PKM), 5 models, estimation sub-sample (140’000 OD)

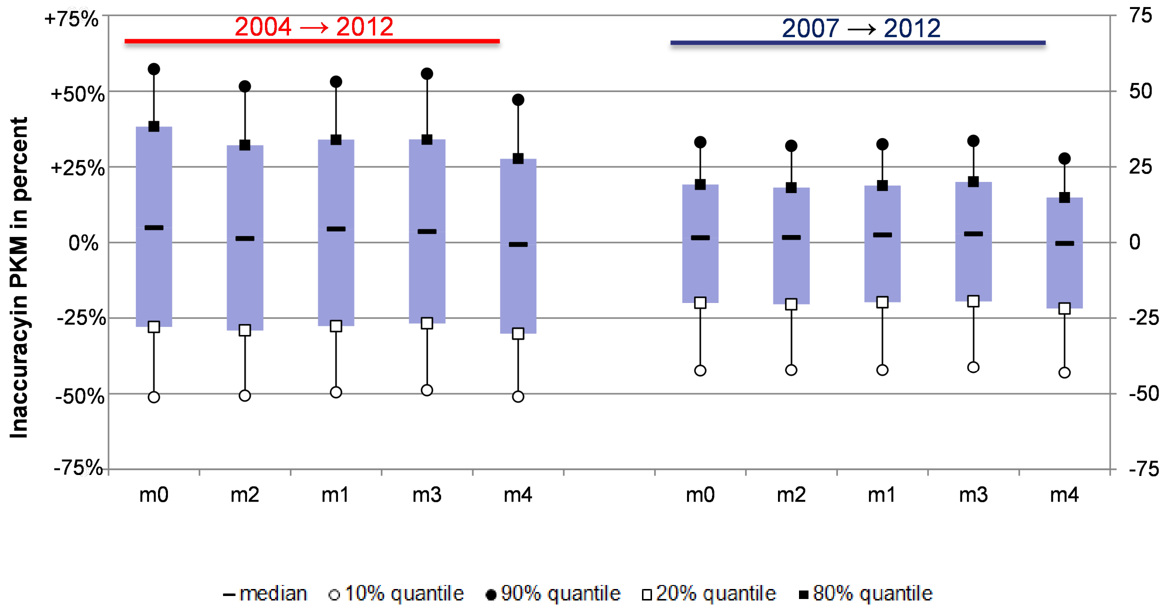
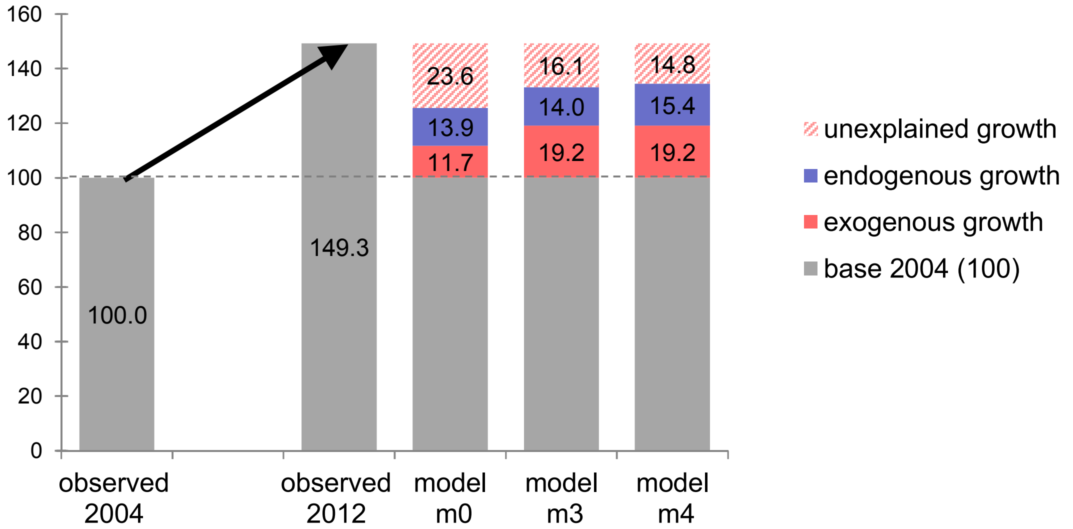


Table 4 Prediction Success – RMSE – extended sample (140’000 OD pairs)

		m0	m1	m2	m3	m4
2004 → 2012	P-Trips	23.9	21.9	22.6	22.0	19.9
2007 → 2012	P-Trips	14.1	13.7	13.8	13.8	12.2
2004 → 2012	PKM	935	828	870	863	730
2007 → 2012	PKM	484	479	530	503	421

We now turn to a macroscopic view of the problem (Figure 7). While the purpose of the model is to predict ridership on the level of individual OD pairs, we show here the aggregated result over the extended OD sample, in comparison of models m0, m3 and m4. The new model m4 explains the observed growth more accurately than the previously used model m0. The exogenous variables have a significantly stronger effect, endogenous variables have a slightly stronger effect and the need to add unexplained growth (UTG) to the prediction is reduced. The need for unexplained growth is the lowest for m4. There is still a need for UTG of 14.8%, which will be discussed in sections 5 and 6.

Figure 7 Prediction of observed growth by groups of variables (P-Trips 2004→2012)



4.4 Prediction success on the full sample

Some OD pairs are not suited to evaluate their specific prediction success, in particular when they were not covered consistently by the onboard surveys in all three years (2004, 2007, 2012). In addition, some OD pairs have a too small number of passenger trips because of insufficient survey representation. To be able to measure prediction success for the entire Swiss rail system, using all 155'000 OD pairs together, we used data aggregation. Two ways of aggregation have been applied: aggregation of passenger demand to flows between districts, and aggregation of network flows by computing route choice for all ODs (assignment).

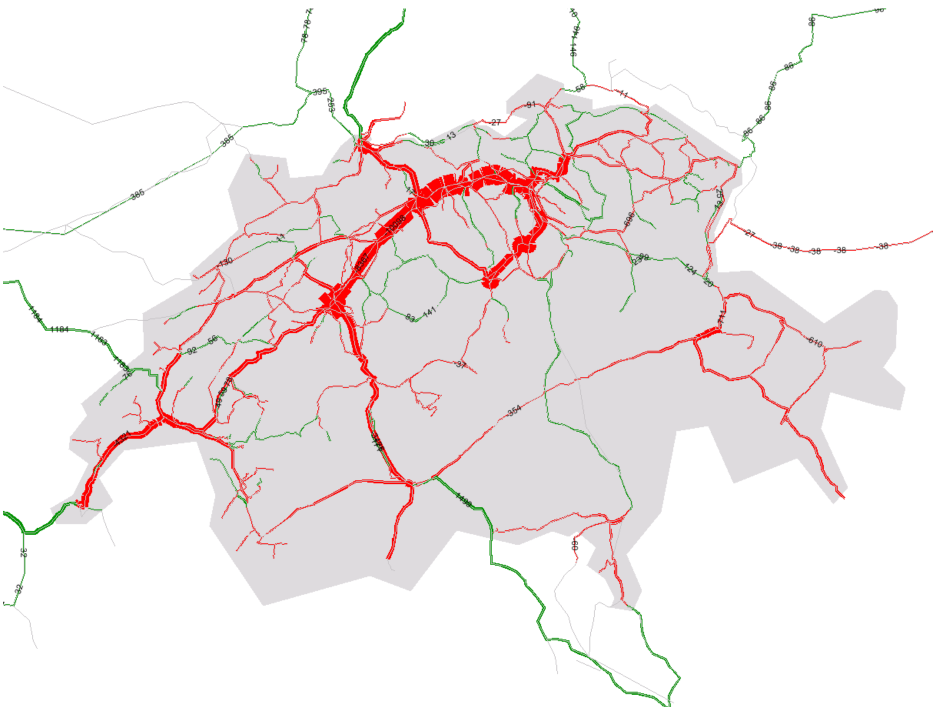
We found that aggregated prediction success statistics provided a better basis to compare the different models, when we restricted the models by taking away the "unexplained growth factor" (which is derived from the constant of the linear regression model). The restricted models' prediction success was interpreted as the models' explanatory power, i.e.: how well can a model predict the future on the basis of all endogenous and exogenous variables, but without constant growth.

Table 5 Minimal absolute error (MAE) of district-to-district flows (P-Trips), restricted model prediction (prediction without unexplained growth), test 2007 → 2012

Best fitting model 2007 → 2012		Agglo Zürich 101	Agglo Basel 102	Agglo Bern 103	Agglo Léman 104	Rest Romandie 120	Ticino 130	Rest Deu-CH 150	Jura 180	Alpen 190
Agglo Zürich	101	**m4**	m4	m2	m3	m3	m2	**m4**	m4	m4
Agglo Basel	102		**m2**	m2	m3	m3	m1	m4	m4	m3
Agglo Bern	103			**m4**	m3	m4	m3	**m2**	m2	m4
Agglo Léman	104				**m4**	m4	m1	m4	m2	m4
Rest Romandie	120					m4	m3	m4	m1	m4
Ticino	130						m4	m1		m4
Rest Deu-CH	150							**m4**	m4	m4
Jura	180								m4	m3
Alpen	190									m4

MAE, minimum over 5 models 2007 → 2012		Agglo Zürich 101	Agglo Basel 102	Agglo Bern 103	Agglo Léman 104	Rest Romandie 120	Ticino 130	Rest Deu-CH 150	Jura 180	Alpen 190
Agglo Zürich	101	-11	-11	-9	-6	-13	1	-8	-3	-3
Agglo Basel	102		-8	-16	-6	-10	6	-7	-8	-9
Agglo Bern	103			-10	-11	-10	-11	-9	-4	-9
Agglo Léman	104				-13	-14		-3	-13	-9
Rest Romandie	120					-16		-7	-19	-4
Ticino	130						-24	-6		-20
Rest Deu-CH	150							-10	-8	-8
Jura	180								-8	
Alpen	190									-14

Figure 8 Prediction success (minimal absolute error): on aggregated assignment results restricted model prediction (prediction without unexplained growth), test 2004 → 2012



Best fitting model	P-Trips	PKM	P-Hours
Regional lines, SBB	m4	m4	m4
Long distance lines, SBB	m4	m3	m3
Total SBB	m4	m4	m4

As can be seen in Table 5, model m4 has the lowest absolute error for 24 of 44 district-to-district pairs, representing 86% of the P-Trips.

The analysis of assignment flows was aggregated by links, train lines, service categories (regional versus long-distance) and sub-networks. This final analysis again confirmed that the new estimated models performed better than the previously used model and that model m4 performed best of all new models. Figure 8 offers a glimpse into this analysis, showing a difference plot between observed passenger volumes and the restricted model's prediction on the network and a table of model comparison for P-Trips and PKM for service categories.

5. Consequences drawn for model improvement

The database covers a time period that is characterized by major changes of all endogenous and exogenous variables. Furthermore, the period is long enough that ramp-up effects (i.e. the 3- to 5-year period until demand has fully reacted to big service changes) can be neglected. Therefore, we consider the elasticity parameters that have been determined with this database as valid and defendable.

Still, these results have not been adopted in SIMBA by relying alone on the statistics shown in section 4. Instead, we went through a further model selection and model validation process, that included practical considerations and a review by service planning practitioners. These practitioners look back on over 15 years of SBB experience with model predictions, where rail demand in Switzerland was predicted reasonably well, especially on an aggregate level. On the other hand, service planners were keen to see if new model parameters would remediate some acknowledged weaknesses of the previously used model.

5.1 Model selection

In this section we explain the process to select the most effective model. The following criteria were considered:

- fit statistics of the linear regression,
- prediction success results,
- balance between simplicity and complexity (we consider simplicity an important goal in model design), and
- usefulness in practice and suitability for application in rail service planning.

In the following we explain how these criteria led to the selection of m4 as the model to be used in SIMBA:

- All the new models show a significantly lower prediction error than the previous model, both in the case of 2004 → 2012 and in the case of 2007 → 2012. Therefore it became clear that the previous model (m0) has to be replaced by one of the new models.
- An important decision was whether the elasticities for TT, AT and NT were to be stratified by OD clusters (example m2) or whether linear-variable elasticities should be used (examples m3, m4). While the cluster-stratified approach performed often better in the estimation fit and in prediction accuracy, we preferred the linear-variable form because of its simplicity, and because of the arbitrary nature of OD clusters as such ("where does agglomeration end and where does intercity start?").
- Simplicity was a major argument to choose a model with linear-variable elasticities. We rejected more complex models with quadratic elasticities and - as explained in the bullet point above - the stratified approach because we felt that they did not reach the perfect balance of simplicity versus complexity.
- From all models m4 was chosen to be the preferred model, because it has the best coefficient of determination (adjusted R^2). It then reaches the best or almost best prediction accuracy and has a better balanced distribution of prediction errors (median close to 0) for P-Trips and PKM. Also, it explains the observed growth to a high degree and hence lowers the need to add unexplained trend growth (UTG).

Another selection criterion which we applied, was the model's usefulness in practice; in other words, if transportation planners at SBB would accept the model. In fact, the discussion of forecasts performed with new parameter sets and the analysis of the results by experienced planners lead us to review the estimation methods and we came up with model m4, after the previous preferred candidate (similar to m3), was criticized for not being plausible in its reaction to service quality changes in the segment of long-distance travel.

5.2 The new model

The results of estimation and prediction success tests lead to the decision to choose model m4 as the future parameter set in SIMBA to forecasting the domestic rail demand. In this section we compare the new model m4 with the one previously used m0 in Table 6.

These new parameters have the following consequences:

- Service quality elasticities are no longer taken as a constant across all trips but are variable as a function of travel time and service frequency in the reference case.

- Service quality elasticities have shifted in their importance: the impact of changes in travel time and service frequency is lowered, the impact of directness (number of transfers) is increased.
- The exogenous variables also have a stronger impact on demand than in previous practice.

Table 6 New versus previously used elasticity parameters

explanatory variable		previous parameter	new parameter
TT	travel time	-1.0	$-0.81 - 0.23 \cdot \frac{TT}{45} + 0.30 \cdot \frac{AT}{15}$
NT	number of transfers	-0.1	$-0.55 + 0.05 \cdot \frac{TT}{45}$
AT	departure adaptation time	-0.4	$-0.33 + 0.08 \cdot \frac{TT}{45} - 0.10 \cdot \frac{AT}{15}$
TA	tarif, rail	-0.4	-0.4
PP	population	+1.0	+1.7
GP	gross domestic product	+0.4	+0.5
RD	travel time, road	+0.6	+0.6
UTG	unexplained growth (p.a.)	[1.01 ; 1.04] - dependent of OD category	

Before applying the new parameters in real forecasting projects, some enhancements of the parameters were still necessary. Mainly we added upper and lower boundaries to the variable elasticity functions, to avoid that service quality elasticities would reach positive values, or even too negative. Another adjustment was a variation of the parameters by trip purpose, which was based on estimation results from other data bases. However the parameters were then calibrated such that the demand reaction of the trip purposes combined would result in the same demand reaction as with the estimated parameters.

As shown in Figure 7, a part of rail ridership growth needs to be represented by UTG, i.e. constant growth rates per OD category. It remains a goal of model development, to further increase the explanatory capacity of the model and to reduce the use of unexplained growth. Still, it is an important feature of a direct demand model, that it can predict high growth rates, while classic four-step models work like zero-sum-game and hence tend to deliver conservative forecasts.

6. Summary and conclusions

We find that longitudinal data are very valuable for model calibration and model validation. As shown in this paper, prediction accuracy of our model SIMBA has been increased as a result of working with longitudinal data. We recommend to travel modeling professionals to dedicate resources to build such data bases and to define prediction accuracy as a major objective of model calibration.

We found evidence that directness (measured as number of transfers) is a strong factor in passenger behavior, and its effect on ridership is stronger than we had previously assumed. On the other hand, the influence of service frequency (measured by adaptation time) is less important – especially for longer trips – than we had previously assumed.

Variable elasticities for service quality have been retained for the use in SBB's travel model SIMBA. Implementation in the model procedure has been completed recently and the application in service planning projects has already started.

The new elasticity parameters have been demonstrated to SBB service planners and the impact of the new parameters on the evaluation of particular rail projects has been tested and discussed. The change of the parameters is now widely accepted.

Current model development at SBB continues to aim for a reduction of UTG (unexplained growth). The goal is to include additional variables in the forecasting model, such as urban density. We will also continue to monitor ridership growth, which has shown signs to slow down since 2012, a fact that gives confidence that the share of unexplained growth in our model will decrease. In the short term, for the next long-term demand scenarios, we will enhance the model to include competition by intercity coach lines, demographic effects, and autonomous vehicles in the forecast. For technological developments such as autonomous vehicles, multi-modal and more detailed modeling approaches are being evaluated.

7. Acknowledgments

The approach of model calibration described in this article has been developed in SBB's travel forecasting practice, with contributions of several members of our modeling team. In particular we would like to thank Marcus Riedi and Andreas Meister for developing the longitudinal database that we used for this project, and for the critical review of the results. Then we would like to thank Johannes Lieberherr for his overall guidance in model development as well as for his help in solving mathematical problems along the way.

8. References

ATOC Association of Train Operating Companies (2013). Passenger Demand Forecasting Handbook. PDFH. Confidential publication reserved for ATOC members. London.

Bützberger, Patrick (2013): Analyse der erklärenden Faktoren für das Wachstum der Bahnnachfrage im Personenverkehr – Verfahren der multiplen Regression zur Verfeinerung der Prognosemethodik im SBB-Modell SIMBA. Master thesis, Universität Bern.

Flyvbjerg, B., Holm, M., Buhl, S. (2005). How (in)accurate are demand forecasts in public works projects? Journal of the American Planning Association, 71(2), 131-146.

Friedrich, M., Hofsäß, I., Wekeck, S. (2001). Timetable-based Transit Assignment Using Branch & Bound Techniques. Transportation Research Records, No. 1752, p. 100-107.

Gentile, G., Florian, M., Hamdouch, Y., Cats, O., Nuzzolo, A. (2016): The Theory of Transit Assignment: Basic Modelling Frameworks, in Gentile, G., Noekel, K. (Eds.) *Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems*, 287-386, Springer, Cham.

Hartgen, D. (2013): Hubris or humility? Accuracy issues for the next 50 years of travel demand modeling. Transportation, 40 (6), 1133-1157.

Kaeslin, L., Lieberherr, J. & Scherr, W. (2014). Demand Data for Dynamic Passenger Assignment within the Swiss National Rail Model. Conference paper, STRC 2014, Ascona.

Lieberherr, J., Pritscher, E. (2012). Capacity-restraint railway transport assignment at SBB Passenger. Conference paper, STRC 2012, Ascona.

Næss, P., Flyvbjerg, B., Buhl, S. (2006). Do road planners produce more 'honest numbers' than rail planners? An analysis of accuracy in road-traffic forecasts in cities versus peripheral regions. *Transport Reviews*, 26 (5), 537-555.

Olesen, A., Bützberger, P., Lieberherr, J. (2016): Modellierung und Bewertung von Fahrplanangeboten der Zukunft. Schweizer Eisenbahn-Revue, 01/2016, S. 16-19.

Quandt, R. E., Baumol, W. J. (1966), "The demand for abstract transport modes: theory and measurement," Journal of Regional Science, 6 (2), 13-26.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Sammer, G., Röschel, G., Gruber, C. (2014): Qualitätssicherung für die Anwendung von Verkehrsnachfragemodellen und Verkehrsprognosen. Bundesministerium für Verkehr, Wien.