Airline Revenue Management as a Control Problem – New Demand Estimates

Philipp Bartke*[1], Catherine Cleophas†[2], and Natalia Kliewer‡[3]

[1] *Department Information Systems,, Freie Universitaet Berlin*
[2] *Research Area Advanced Analytics,, RWTH Aachen University*
[3] *Department Information Systems,, Freie Universitaet Berlin*

November 19, 2013

Abstract

Revenue optimization is usually based on a model of market demand. In practice, the true model cannot be known and has to be estimated from observed sales and availabilities. As airline revenue management systems become increasingly sophisticated, the number of parameters in the demand model grows, rendering demand estimation a challenging endeavor.

This paper formulates the demand estimation and revenue optimization problem as a state-space model and illustrates that it closely relates to well-known models in control theory. Based on this, we apply techniques developed in this field for Bayesian Learning and Dual Control, Kalman and Particle Filters.

In a simulation study, we evaluate these methods as adapted to demand estimation for revenue management using the Posterior Cramér-Rao Bound as a benchmark. We compare the results to those achieved by two more common approaches, simple estimation and maximum likelihood estimation.

*`philipp.bartke@fu-berlin.de`
†`catherine.cleophas@rwth-aachen.de`; Corresponding author
‡`natalia.kliewer@fu-berlin.de`

With respect to forecast quality, we find the performance of the Unscented Kalman Filter to be superior to the alternatives considered. In addition, we point out that this estimation method yields auxiliary information about the uncertainty of the current estimate from which estimation itself, but also revenue optimization and interaction with a human analyst may benefit.

1 Introduction

Airline revenue management (ARM) as a field of operations research has been focused on maximizing revenue through optimal inventory controls at least since the publication of Littlewood (1972). Earlier research concentrated on the problem of overbooking, preparing the way by establishing a feedback loop between observations, forecast, and capacity allocation. Based on a forecast of differentiated demand segments, the same product can be offered at multiple prices, thereby optimally utilizing the different customers' willingness-to-pay. As the segmentation of demand follows a forecast that is based on demand estimates derived from observed sales, the quality of these estimates often decisively influences ARM success. For a short account of the history of revenue management with special regards to the airline industry please refer to Horner (2000); a thorough introduction to established methods is given in Talluri and van Ryzin (2005).

As pointed out in the literature review of this paper (Section 2), classic operations research approaches to ARM frequently regard forecast and optimization as independent problems. So far, very few contributions applying control theory to revenue management seem to exist. By considering ARM from the perspective of control theory, we emphasize the consequences of the underlying feedback loop: Sales are observed based on inventory controls that are optimized according to demand estimates derived from sales.

Our contribution aims to further the application of control theory to ARM by introducing a state-space model and using it to introduce new demand estimates based on Unscented Kalman and Particle Filters (UKF and PF). We evaluate the results using the Posterior Cramér-Rao Bound (PCRB) as a benchmark. Comparing the results to those of simple estimation (SE) and maximum likelihood estimation (MLE), we find that UKF and PF produce a forecast error that is orders of magnitude smaller than that of SE and performs close to the MLE and the PCRB. Moreover, lost revenue compared to what could be achieved given so-called "perfect" forecast is less than 1%. As detailed in Section 6, not only do approaches to demand estimation as derived from control theory perform well for revenue management, they also provide information on the uncertainty of the current estimate. This additional information is potentially valuable for finding the appropriate level of detail in forecasting, improving optimization and helps the human analyst reconciling her own intuition with demand estimates from the ARM system.

The following section first summarizes research on revenue management that explicitly considers the problem of demand learning. It introduces literature on the methods we adapt (UKF and PF) and the benchmark we evaluate them against (PCRB). Following the description of the ARM process from a control perspective (Section 3), we formulate demand estimation and revenue optimization as a state-space model (Section 4). Based on this, we adapt techniques developed for Bayesian Learning and Dual Control. In a simulation study, we evaluate the results with respect to revenue and forecast quality, comparing them to simple estimation and maximum likelihood estimation (Section 5). The final section discusses our contributions in further detail and provides an outlook to future research.

2 Literature Review

This section introduces relevant research as the basis for the proposed model and approaches. First, existing efforts combining revenue optimization and demand learning are summarized and related to our contribution. The second part of this review provides an introduction to Kalman and Particle Filters as tools for estimation. Finally, we introduce the Posterior Cramér-Rao Bound as a benchmark for demand estimates.

2.1 Revenue Management with Demand Learning

Many works on revenue management assume that the underlying demand model is known and exclusively focus on improving optimization methodology. However, a small stream of research examining dynamic pricing problems is closely related to our work in that it considers the problem of demand learning explicitly.

Bitran and Wadhwa (1996) consider a dynamic pricing problem for seasonal products. They model customer arrivals as a Poisson process with a known but potentially time-variant rate. Each customer's reservation price is drawn from a probability distribution with a potentially unknown parameter and may vary over time. There is only a single product, which a customer will purchase if her willingness-to-pay exceeds the price of the product. Bitran and Wadhwa (1996) develop a Bayesian update procedure for the reservation price distribution parameter, assuming that the arrival rate is known. They allow for some demand changes between time periods, but these cannot be random and have to be known to the modeler.

Lobo and Boyd (2003) also consider a dynamic pricing problem, using a linear demand model with an intercept and one coefficient corresponding to price. The parameters of the model are unknown and drawn from a Gaussian distribution. The authors provide equations of the Bayesian update that are equivalent to Kalman Filter equations. In this setting, they consider the active learning problem. In traditional learning, demand estimation is passive in that it only observes prices or availabilities and cannot directly influence them. However, an incentive to select prices or availabilities that improve demand estimation can exist. This will result in short-term revenue losses, but these may be outweighed by future revenue gains based on superior demand estimates. Lobo and Boyd (2003) develop an approximate solution to the active learning problem, using convex semi-definite programming techniques. Our work is closely related to theirs, but focuses on revenue management rather than on dynamic pricing. Also, we allow for additional flexibility in the specification of the demand model.

Carvalho and Puterman (2004) assume a log-linear demand function with unknown parameters and also rely on Kalman Filter equations for Bayesian Learning. The authors develop a one-step-look-ahead strategy based on a second degree Taylor expansion of the expected future revenue as a heuristic solution to the active learning problem. They compare this pricing policy to various other schemes, including a myopic strategy, random price variation and a "softmax" strategy using Monte Carlo simulations. In their setting, the myopic policy clearly under-performs, and the one-step-look-ahead strategy yields slightly higher expected revenues than the alternatives.

Aviv and Pazgal (2005) provide a model of Bayesian Demand Learning where customers arrive in a Poisson process with unknown rate. They model arrival rate uncertainty as a Gamma distribution to achieve a simple update rule for the belief distribution. Price-sensitivity is modeled as an exponential distribution with a known mean. The authors focus on the distinction between active and passive learning. In their setting, the benefits of active learning are minor as long as the level of uncertainty is not too high (this contrasts with the results in Carvalho and Puterman (2004)). Aviv and Pazgal (2005) conclude that the passive learning approach is a reasonable heuristic. Intuitively, this may be attributed to the assumption that price-sensitivity is known.

Sen and Zhang (2009) also model demand as a Poisson process with an unknown arrival rate. The reservation price distribution is unknown as well, but can be derived from a finite set of candidate distributions. The authors provide a Bayesian Learning model to jointly estimate the arrival rate and the reservation price distribution. The information requirements of their method increase with the number of candidate distributions, making it crucial for practical implementations to restrict the candidate set as much as possible.

Vulcano et al. (2012) consider a model where customer arrivals form a Poisson process and customers choose from the subset of currently available products according to an multi-nomial-logit model of choice. They present a re-formulation of the estimation problem in terms of so-called "Primary Demand": the demand for a product when all alternatives are available. The re-formulation yields a much simplified estimation maximization (EM) procedure to find a maximum likelihood estimate, both of the arrival rate and the product valuations. The method developed is not a demand learning method per se, since it does not allow for incremental updates in the demand estimates. Instead, all historical data has to be processed every time the demand estimate is updated.

A slightly different approach is taken by Stefanescu (2009) and Kwon et al. (2009). Stefanescu (2009) models demand as a multivariate Gaussian distribution. She argues that customer choice modeling may not be appropriate in the face of customer heterogeneity or missing data, e.g. about competitors' offers. Her model uses demand correlation to account for time and inter-product dependence. The author develops an EM algorithm to estimate this model based on censored data, which shows promising results. Again, this is not strictly demand learning, since the EM algorithm requires the complete data set to update the current estimate. Moreover, the descriptive nature of this demand model seems less suited as an input for revenue optimization.

Kwon et al. (2009) consider "non-cooperative competition among revenue maximizing service providers". Each firm uses a Kalman Filter to estimate the parameters of the demand model. Demand is deterministic and independent between different products, but it depends exclusively on past and current market prices in a linear fashion. The authors assume that the coefficients of their demand model evolve according to a random walk, similar to the assumption in this thesis. However, they model the dynamics of demand parameters only over a single, continuous selling horizon. While this model seems appropriate in a retail setting, it does not realistically capture demand dynamics in airline revenue management, where demand evolves both over the selling horizon of a particular flight and between consecutive flights.

Li et al. (2009) and Chung et al. (2012) extend the model of Kwon et al. (2009) by allowing for a much more general form of demand evolution over the selling horizon. Moreover, they highlight the notion of a state-space model to formulate the dynamic pricing and demand estimation problem and use a Markov chain Monte Carlo technique for parameter estimation. Yet, their demand model is still very limited, in that it does not include stochastic demand, dependence between products nor demand evolution between consecutive flights.

In this paper, we consider the ARM problem for a single firm, formulating it as a state-space model. In contrast to the group of papers mentioned above Kwon et al. (2009); Li et al. (2009); Chung et al. (2012), we do allow for stochastic demand and a more general, non-linear price-dependence of demand. Moreover, demand evolution and demand learning is accomplished between consecutive selling seasons and not within. As such, we believe that our model is more suited in the context of *airline* revenue management. Based on our model, we adapt non-linear Kalman and Particle Filters to create new estimates from censored data. In contrast to the method developed in Vulcano et al. (2012), the filter methods are applicable to a wider range of demand or choice models and completely avoid the convergence problems of the EM algorithm.

We are not the first to use a Kalman filter for demand estimation in a revenue management context, as noted in the literature review. In fact, (Talluri and van Ryzin, 2005, p. 458ff.) describe the Kalman Filter as a general method for time-series forecasting, however without making the actual connection to a concrete revenue management problem. In this paper, we pick up from there and show how to model the airline revenue management problem such, that it is amenable to the general Kalman Filter formulation. We observe that the assumptions of the original Kalman Filter are too strict for most demand models, and consequently we consider extensions of the Kalman Filter to non-linear and non-Gaussian models.

2.2 Kalman Filter and Extensions

The Kalman Filter has been designed to iteratively estimate the hidden state of a system based on indirect and noisy observations. Starting with the original paper of Kalman (1960), a tremendous amount of work has been put into extending the model in various directions and adapting it to numerous application areas. We limit our overview to contributions that are of particular relevance to the work presented in this paper.

Kalman (1960) considers a state-space model, in which the state of a system at a time t is a linear function of the state at time $t-1$ plus some Gaussian error term. This state, however, is never directly observed; instead, it is only available through a linear observation function with some Gaussian measurement noise. Both the state evolution function and the measurement function can be time-dependent. Given a Gaussian prior distribution for the system state, one can show that both the posterior as well as the prior for the next time step are also Gaussian and that the minimum least squares estimator has a closed-form solution. Moreover, past observations influence the current estimate only through the prior distribution of the system state. As this distribution is Gaussian, it is sufficient to keep track of the mean and the covariance matrix of the system state: old observations can be discarded. These properties make the Kalman Filter computationally efficient in terms of time and memory requirements.

The original Kalman Filter formulation puts strong assumptions on the underlying process model. In our model, we identify the observation function with the demand model. Hence, requiring a linear observation function would severely restrict the set of permissible demand models. Additionally, the original model assumes observations to also follow a Gaussian distribution. Even though demand has been modeled as such in ARM literature (Belobaba (1989)), this assumption seems to be inappropriate for booking data that is non-negative and integer, especially if overall demand is relatively small.

Julier and Uhlmann (1997) address the problem of non-linearity in the state evolution function and in the observation function. They use an approximate approach based on an alternative parametrization of the normal distribution, the so-called unscented transform, resulting in an Unscented Kalman Filter (UKF). While still conceptually simple and computationally efficient, this method outperforms other methods overcoming the linearity restriction. We therefore selected the UKF as one of our candidate methods and adapted it to our problem domain in which the form of the observation function enables significant computational short-cuts.

2.3 Particle Filter

The Particle Filter (PF) represents an alternative approach to the above described Kalman Filters. It employs Monte Carlo sampling to estimate the state of a system; in our case, the ARM demand model. PF uses a large set of discrete points – a “particle cloud” – to approximate the posterior probability distribution. As such it requires no assumption on the parametric form of this distribution. PF is closely related to Monte Carlo integration methods with importance sampling. Therefore, it is also known as a Sequential Importance Resampling Filter.

The earliest reference to this approach that we are aware of is Müller (1991), who proposes a PF with rejection sampling to estimate the parameters in general dynamic models. Two years later, Gordon et al. (1993) introduce an Importance Resampling Filter, which is very close to the method presented here. However, while Gordon et al. (1993) perform the resampling at every time step, we only resample when necessary. Independently of these two references, Kitagawa (1996) proposes essentially the same algorithm as Gordon et al. (1993). Doucet et al. (2000) provide a review of a variety of PF methods and develop a general framework.

2.4 Posterior Cramér-Rao Bound

Finally, we turn to a concept for the evaluation of estimates: The Cramér-Rao bound provides a lower bound for the mean squared error of an estimate, which has to hold for any concrete estimation method.

The original Cramér-Rao bound is based on time-invariant models. However, an extension, the Posterior Cramér-Rao bound (PCRB) is applicable in the context of this paper. Its dynamics over time for the discrete-time nonlinear filter problem were derived by Tichavsky et al. (1998).

This bound can be computed for many models where an exact solution to the estimation problem is not available. It can thus serve as an absolute benchmark to compare approximate estimation methods against – such as UKF and PF described above.

3 Airline Revenue Management from a Control Perspective

A control perspective highlights the ARM feedback loop linking inventory optimization and sales. This emphasizes the resulting complexity of demand estimation from censored data. The following description of the ARM model from a control perspective creates the foundation for the state space model and the adaptation of filtering methods documented in the next section.

3.1 Airline Revenue Management Model

We consider an ARM system with distinct estimation and forecasting. The system is abstract in that it assumes no specific methods for demand estimation, forecasting, optimization or inventory control. We assume that the system can be decomposed into forecaster, optimizer and inventory and that the system controls the availability of booking classes rather than controlling the price directly.

ARM methods without distinct forecasts, such as reinforcement learning and a large body of literature on dynamic pricing violates the above assumptions. Nevertheless, in the airline industry, the additional value of demand forecasts extends beyond revenue management, e.g. forming an input to fleet assignment. The pervasive use of the booking class standard throughout the reservation and check-in processes prevents the adoption of dynamic pricing methods in practice in the foreseeable future. Therefore, we believe the model presented here is general enough to describe the ARM systems used by traditional network carriers.

Figure 1 provides an overview of the ARM system as a whole. In the real world, the "Update" step is performed by deriving new bookings from the interplay of inventory controls with market demand. In contrast, a simulation system as used in Section 5 relies on artificially generated demand. The remainder of this section describes the system components in further detail.

3.1.1 Forecaster

The forecaster is defined by a demand model and an estimation procedure for that model. The demand model provides a mapping H_a from product availabilities a and a vector of demand parameters x to a distribution of bookings for each product $H_a(x)$. Here, a product is any individually sold

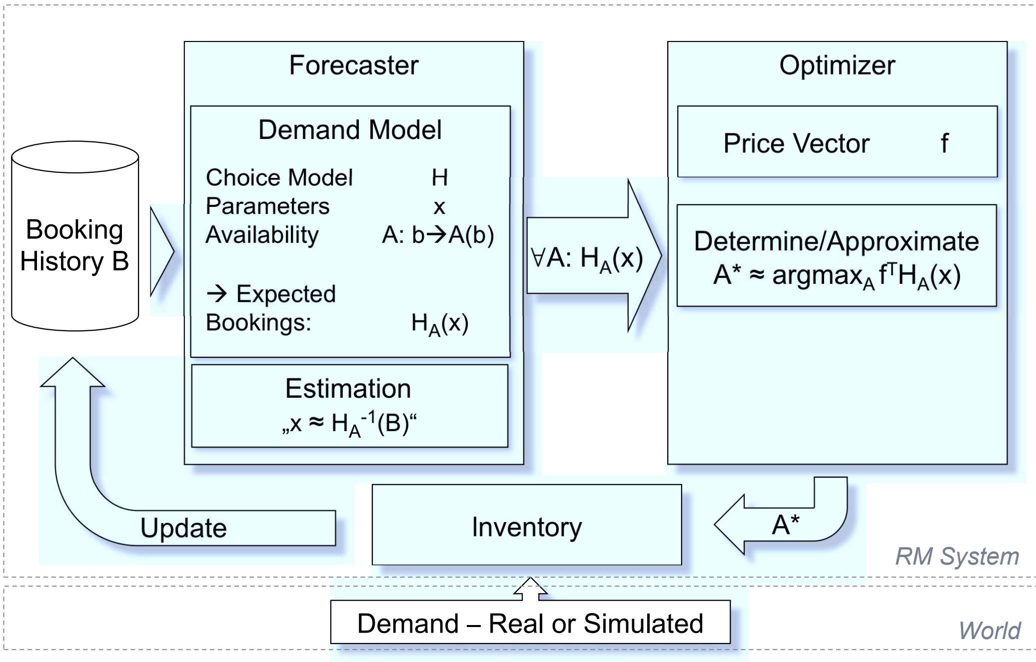

Figure 1: Our model of a traditional ARM system

offer, typically a specific booking class for a specific itinerary over the airline's network on a particular date and time, sold a specific number of days before departure in a specific point of sale. The availability of a product usually depends on the number of seats already sold or on the flights included in the itinerary. Hence, we express the availability as a function of the current vector of bookings b. Capacity is assumed to be constant and is therefore suppressed from our notation.

The choice model $H_a(x)$ can be any function of a and x, as long as the restrictions on permitted sales for each product as prescribed by a are obeyed. E.g., $H_a(x)$ could assume independent demand, in which case expected bookings for a product only depend on the product's own availability. In the most general case, expected bookings for a product depend on the availabilities of all other products. Independent demand is an unrealistic assumption but has been used in practice due to its simplicity. The completely dependent model provides the most flexible description of demand, but with billions of offered products, it quickly becomes intractable. Realistic and efficient demand models therefore severely restrict the set of dependencies, e.g. to the set of booking classes for the same itinerary or to a set of similar itineraries for the same combination of origin and destination.

The parameter vector x encapsulates whatever is a priori unknown about the demand model. The length of the vector and the interpretation of its components depends on the concrete model. When assuming independent demand, x may simply represent the expected bookings per available product. For dependent demand, at least some components of x have to describe the dependency on other products' availabilities. This may be implemented, for instance, by including a price-sensitivity parameter in x or by modeling overlapping demand. Additionally, x may also contain seasonality factors or weekday patterns.

Since x is unknown to the modeler, it has to be estimated from observed sales. This means that we try to find the probability distribution of x conditional on the history of booking vectors $B = \{b_1, \ldots, b_T\}$ and the history of availabilities $A = \{a_1, \ldots, a_T\}$. Alternatively, we might only obtain some property of this distribution, such as its mode, which would yield the maximum-likelihood estimate, or its mean, which would yield the expected value of x conditional on the observations. We defer further discussion of the estimation problem to Section 3.2 and continue with the assumption that there exists some estimate of x, either a probability distribution or a point-estimate.

Finally, the forecaster computes the expected bookings (or the distribution thereof) for all feasible availability functions using the current demand estimate x. Availability functions are feasible if they obey the capacity constraints and can be implemented by the inventory. The predictive power of the demand model is crucial for this step, as it allows the forecaster to provide the number of expected bookings even for availability situations that were never observed. In practice, no forecaster iterates over all feasible availability functions and computes a list of expected booking vectors: Such a list would be prohibitively long. However, characteristics of efficient demand models such as limited dependence usually allows for a much more efficient encoding. Moreover, the optimization method's requirements may also limit the amount of information relevant. Information that is not used by the optimizer need obviously not be computed by the forecaster.

3.1.2 Optimizer

The optimizer combines the expected bookings $H_a(x)$ for all a received from the forecaster with the vector of prices of each product f. It calculates the availability a^* that maximizes $f^T H_{a^*}(x)$. This conceptually simple step can usually not be solved to optimality in practice due to the very large number of potential availability functions a. Hence, heuristic methods find an approximate solution to the optimization problem.

The optimizer's output is the optimal availability function a^*. This function is sent to the inventory to control the acceptance of bookings.

3.1.3 Inventory

The inventory implements the availability function a^*: It offers all products, for which $a^*(b)$ is true and prevents the sale of all other products, b being the vector of current bookings. Actual inventory systems constrain the set of availability functions that can be implemented. Inventory systems may constrain the number of bookings in a certain booking class on a particular flight by booking limits or protection limits. Alternatively, they can set a bid-price for each flight and only make products available for which the price exceeds the sum of bid-prices of the flights in the itinerary. Combinations of these techniques are also possible.

After some time interval, the booking history B and the availability history A are updated and augmented by new observations. This triggers a new loop through the complete ARM process, from estimation to prediction to optimization to implementation.

3.1.4 Feedback Loop

The fact that ARM systems include a feedback loop renders their long-term dynamics non-trivial. Even for relatively simple components, the overall behavior may be complicated and hard to predict. Choosing an availability in the current time-period will influence the observations in the next time-period, which will influence the future demand estimate. This may in turn lead to a revised availability.

Specifically, if something akin to price-sensitivity is to be estimated, it may well be worth to occasionally set availabilities that are not short-term optimal in order to learn more about the sensitivity. If the short-term optimal availability is implemented at all times, no new information about the sensitivity parameter may be gained. In this case, the system might get stuck in a state far from the actual optimum.

In a realistic ARM system, the analysis of the feedback loop may be close to impossible. Yet, the overall performance of the system strongly depends on the feedback dynamics and a system of components that ignores this feedback loop may be far from optimal.

3.2 Estimation Problem

For the estimation problem, we assume that the real demand parameters are the realizations of an auto-regressive process of order 1 (AR(1)-process). Auto-regressive processes are frequently used to model various kinds of time series, including demand time series. We restrict ourselves to the following particular form:

$$x_{t+1} = x_t + w_t \quad w_t \sim N(0, Q) \tag{1}$$

In other words, the change of parameters from time t to time $t+1$ is described by a multi-variate Gaussian random variable w_t with zero mean and covariance matrix Q. If the range of some parameters has to be constrained to provide meaningful inputs for the choice model, we have to assume a truncated normal distribution for w_t. In the following discussion, we assume that the real parameter values are far enough from the bounds of the valid range for a truncation of the normal distribution to have only a minor effect. Accordingly, we neglect this aspect.

Further, we assume that bookings follow a Poisson distribution conditional on their mean. This distribution can be censored if the availability function a limits the number of available seats. Since usually, many products compete for the same capacity and a majority requires more than one seat, this censoring may link the booking distribution of almost all products in a non-trivial way. This is the case even if the demand model does not include dependencies.

We use the following heuristic to approximate the situation: We assume that we know the fraction s_i of the observation period when product i was available. Then, we let the bookings for product i be Poisson distributed with an arrival rate of $s_i \cdot \lambda_i$.

4 Improving Estimation for Airline Revenue Management

The framework introduced in Section 3 lends itself well to filter-type estimation methods. Specifically, we have the following state-space model:

$$x_{t+1} = x_t + w_t \qquad w_t \sim N(0, Q) \tag{2}$$
$$b_t \sim Poi(H_a(x_t)) \tag{3}$$

where, again x_t is the vector of demand parameters at time t, b_t is the vector of observed bookings at time t and H_a is the choice function for availability a.

If H_a was a linear function and the Poisson distribution in equation 3 is approximated by additive Gaussian noise, we would arrive at a standard linear state-space model with additive Gaussian noise:

$$x_{t+1} = x_t + w_t \qquad w_t \sim N(0, Q) \tag{4}$$
$$b_t = H_a x_t + v_t \qquad v_t \sim N(0, R) \tag{5}$$

Due to its restrictive assumptions this is not a suitable model for our problem. However, this model is instructive to see the connection to the Kalman Filter, since it is the standard estimator for x_t in this type of model Kalman (1960). The Kalman Filter has a number of desirable properties. It

- is the minimum mean squared error (MMSE) estimator and attains the PCRB exactly;
- is computationally fast;
- can be computed recursively, that is only the last estimate $\hat{x}_{t-1}$, its covariance P_{t-1} and the current observation b_t are required to produce the next estimate $\hat{x}_t$ and its covariance P_t;
- produces not only a point estimate, but the complete posterior distribution of x_t given the observation history by providing the covariance of the estimate P_t.[1]

[1] This may be useful in optimization, e.g. to explore the state-space in the direction of greatest uncertainty.

Due to its advantages, a wide array of applications use the Kalman Filter. However, it is by itself not applicable to the original model from equations 2-3. We therefore present two extensions of the Kalman Filter rendering it applicable while preserving at least some of its desirable properties.

4.1 Unscented Kalman Filter

The Unscented Kalman Filter (UKF, Julier and Uhlmann (1997)) can handle non-linearities both in the state evolution equation and in the observation equation. We only make use of the latter property as the state evolution equation (eqn. 2) of the original model is already linear. As the UKF is a heuristic, it is not necessarily the MMSE estimator. The other properties of the Kalman Filter are preserved, with the caveat that the posterior distribution used is only an approximation.

We formulate the UKF for a "hybrid" choice function that can be decomposed into a linear and a multiplicative non-linear part:

$$H_a(x) = L_a x^L + \lambda C_a(x^N) \tag{6}$$

where $x = (x^L, x^N)$. Isolating the linear part of H_a allows for a much more efficient implementation by combining a regular Kalman Filter for the linear part with the UKF for the non-linear part. If the choice function cannot be isolated, we can simply set $L_a = 0$. If $C_a(x^N) = 0$, our formulation is identical to the original Kalman Filter.

The UKF assumes Gaussian observation errors that are not necessarily additive. Thus, we implicitly assume the following observation equation:

$$b_t = L_a x_t^L + \lambda C_a(x_t^N) + v_t \qquad v_t \sim N(0, diag(H_a(x_t))) \tag{7}$$

where $diag(H_a(x))$ is the matrix with $H_a(x_t)$ on its diagonal and all other entries equalling zero. Thus, the variance of the observation equals its expectation, as is consistent with a Poisson distribution.

At every time step t, the UKF will produce an estimate $\hat{x}_t$ and a covariance matrix P_t that together define the approximate current belief about the real demand parameters. In other words, $\hat{x}_t$ and P_t are the parameters of the approximate posterior distribution of x_t, given all observations up to time t.

The estimation algorithm[2] is as follows. First, decompose the covariance matrix P_t into an upper triangular matrix U, such that $UU^T = P_t$. From

[2]The derivation from the original UKF formulation is straight-forward, exploiting linearity and the upper triangular form of U whenever possible.

this, compute the set of $2n^N + 1$ sigma points σ_i, where n^N is the number of demand parameters of the non-linear part of the choice function:

$$\sigma_0 = \hat{x}_t \tag{8}$$

$$\sigma_i = \hat{x}_t + \sqrt{n+\gamma}\, U_i \qquad i = 1, \ldots, n^N \tag{9}$$

$$\sigma_i = \hat{x}_t - \sqrt{n+\gamma}\, U_{i-n^N} \qquad i = n^N + 1, \ldots, 2n^N \tag{10}$$

where U_i is the i-th column of U, n is the number of demand parameters for the complete choice function and γ is the Lambda-parameter from Julier and Uhlmann (1997), renamed here to avoid confusion with the demand volume parameter.

Next, apply the non-linear part of the choice function to each sigma point to obtain

$$g_i = C_a(\sigma_i) \qquad i = 0, \ldots, 2n^N \tag{11}$$

From this, compute the expected bookings from the non-linear part z^N, the linear part z^L and their sum z:

$$z^N = \frac{\gamma + n^L}{\gamma + n} \cdot g_0 + \frac{1}{2(n+\gamma)} \cdot \sum_{i=1}^{2n^N} g_i \tag{12}$$

$$z^L = L_a \hat{x}_t^L \tag{13}$$

$$z = z^L + z^N \tag{14}$$

Additionally, find the booking covariance matrix from the non-linear part

$$\begin{aligned} P_{z^N z^N} = {} & \left(\frac{\gamma + n^L}{\gamma + n} + (1 - \alpha^2 + \beta) \right) \cdot (g_0 - z^N)(g_0 - z^N)^T \\ & + \frac{1}{2(n+\gamma)} \cdot \sum_{i=1}^{2n^N} (g_i - z^N)(g_i - z^N)^T \end{aligned} \tag{15}$$

and the cross-covariance between the complete state x_t and the non-linear bookings z^N

$$P_{xz^N} = \frac{1}{2(n+\gamma)} \cdot \sum_{i=1}^{2n^N} (\sigma_i - \hat{x}_t)(g_i - z^N)^T. \tag{16}$$

Based on this, and the upper left $n^L \times n^L$ block of P_t P^L, compute the total booking covariance

$$P_{zz} = L_a P^L L_a^T + P_{z^N z^N} + diag(H_a(\hat{x}_t)) \tag{17}$$

and with the left $n \times n^L$ block of P_t P^{NL}, construct the total cross-covariance

$$P_{xz} = P^{NL} L_a^T + P_{xz^N} \tag{18}$$

The values obtained for z, P_{zz} and P_{xz} are now used for the regular Kalman Filter update and prediction equations. Hence, we compute the Kalman gain $K = P_{xz} P_{zz}^{-1}$ and from that the new demand estimate

$$\hat{x}_{t+1} = \hat{x}_t + K \cdot (b_t - z) \tag{19}$$

$$P_{t+1} = P_t - K P_{zz} K^T + Q \tag{20}$$

4.2 Particle Filter

The Particle Filter (PF) is a more general extension of the Kalman Filter, in that it can handle almost any state-space model. This flexibility comes with additional computational effort. The PF starts with a large number of hypotheses (particles) about the real parameters and updates the likelihood of these hypotheses as new observations arrive. The number of particles treated makes the approach computationally more demanding. Conceptually however, the method is very straight-forward and easy to implement.

When the number of particles tends to infinity, the Particle Filter is asymptotically the minimum mean squared error filter and the approximated posterior density converges to the real posterior density (Gordon et al. (1993)). However, for a finite number of particles, little can be said about the quality of the filter; the required number of particles can only be found experimentally. In practice, there exists a trade-off between filter quality and the required computational resources.

Since the Particle Filter puts no restrictions on the state-space model, we use our original formulation as follows:

$$x_{t+1} = x_t + w_t \qquad w_t \sim N(0, Q) \tag{21}$$

$$b_t \sim Poi(H_a(x_t)) \tag{22}$$

Let N be the number of particles used. At every time step, the Particle Filter holds a set of particles $P_t = \{\hat{x}_{tk}, k = 1, \ldots, N\}$ and a corresponding set of weights $W_t = \{\omega_{tk}, k = 1, \dots, N\}$. Each particle represents a potential parameter vector for the choice function H. Its corresponding weight is the likelihood of that parameter vector being the true parameters.

When a new set of observed bookings b_t and availabilities a_t arrives, a new set of particles is generated and their likelihoods are evaluated. The new set could theoretically be drawn from a uniform distribution over the whole parameter space. This, however, would lead to large number of particles with very small likelihoods, and therefore excessive computational requirements. Importance sampling puts most particles in regions of high interest. These regions are described by a so-called importance function $\pi(x|x_{0:t-1,k}, b_{0:t})$, which assigns a weight to each point in the parameter space based on the particle's past trajectory and the observation history. There are multiple choices for the importance function, and most PF variants found in the literature differ primarily in this (Doucet et al. (2000)).

The general algorithm consists of the following steps.

- For $i = 1, \dots, N$ sample $\hat{x}_{tk} \sim \pi(x|\hat{x}_{0:t-1,k}, b_{0:t})$.
- For $i = 1, \dots, N$ compute importance weights:

$$\omega'_{t,k} = \omega_{t-1,k} \cdot \frac{p_a(b_t|\hat{x}_{tk}) \cdot p(\hat{x}_{tk}|\hat{x}_{t-1,k})}{\pi(\hat{x}_{tk}|\hat{x}_{0:t-1,k}, b_{0:t})} \tag{23}$$

- For $i = 1, \dots, N$ normalize importance weights $\omega_{t,k} = \frac{\omega'_{t,k}}{\sum \omega'_{t,k}}$.

The conditional probability $p_a(b_t|\hat{x}_{tk})$ is given by the Poisson probability distribution function and the choice function H:

$$p_{a_t}(b_t|\hat{x}_{tk}) = \prod_i \frac{(h_{a,i}(\hat{x}_{tk}))^{b_{t,i}}}{b_{t,i}!} \cdot e^{(h_{a,i}(\hat{x}_{tk})} \tag{24}$$

The conditional probability $p(\hat{x}_{tk}|\hat{x}_{t-1,k})$ describes the state evolution. From equation 21, we find that this is a multivariate Gaussian distribution with mean $\hat{x}_{t-1,k}$ and covariance Q.

Together, weights and particles form a discrete distribution approximating the actual continuous posterior distribution. The expected value for the parameter estimate at time t can be computed as the mean of the particles at time t, such that $\hat{x}_t = \frac{1}{N} \sum \omega_{tk} \hat{x}_{tk}$.

As mentioned above, it is desirable to keep the particle weights as evenly distributed as possible. This can be measured by the particle weights' variance. It can be shown that the importance function $\pi(x|x_{0:t-1,k}, b_{0:t}) = p_a(x|x_{t-1,k}, b_t)$ minimizes the variance of the particle weights Doucet et al. (2000), which is therefore called the optimal importance function. In our model, this function has no closed form such that analytical sampling from this function is not possible. As proposed in Doucet et al. (2000), we approximate the importance function locally around x by a multivariate Gaussian distribution. We define the log-likelihood $l(x) = \log p_a(x|x_{t-1}, b_t)$ and then compute its first two derivatives:

$$\begin{aligned} l(x) &= \text{const.} - \frac{1}{2}(x - x_{t-1})^T Q^{-1}(x - x_{t-1}) \\ &\quad + \sum_{i:h_i(x)>0} (b_{t,i} \log h_i(x) - h_i(x)) \end{aligned} \tag{25}$$

$$\nabla_x l(x) = -Q^{-1}(x - x_{t-1}) + \sum_{i:h_i(x)>0} \nabla_x h_i(x) \cdot (\frac{b_{t,i}}{h_i(x)} - 1) \tag{26}$$

$$\begin{aligned} \Delta_x^x l(x) &= -Q^{-1} \\ &\quad + \sum_{i:h_i(x)>0} \left(\Delta_x^x h_i(x) \cdot (\frac{b_{t,i}}{h_i(x)} - 1) - (\nabla_x h_i(x))(\nabla_x h_i(x))^T \cdot \frac{b_{t,i}}{h_i^2(x)} \right) \end{aligned} \tag{27}$$

where $\Delta_y^x = (\nabla_x)(\nabla_y)^T = (\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n})^T(\frac{\partial}{\partial y_1}, \ldots, \frac{\partial}{\partial y_n})$.

A second order Taylor expansion yields the covariance $\Sigma = -l''(x)^{-1}$ and mean $m = x + \Sigma \cdot l'(x)$. The point x, around which the log-likelihood function is approximated locally, should be the mode of $p_a(x|x_{t-1}, b_t)$, which can be found numerically with Newton's iterative method. Constructing this importance function is computationally expensive when the number of parameters becomes large, but it has the advantage of putting more weight on areas of the parameter space that are in agreement with the current observation.

The particle cloud for such a filter will still degenerate at some point, as analytically proven in Doucet et al. (2000). That is, most of the particle weight will be concentrated on a single particle. To overcome this, the particles have to be resampled from time to time. We employ the resampling strategy presented in Doucet et al. (2000). If the estimated number of effective particles $N_{\text{eff}} = \frac{1}{\sum \omega_{tk}^2}$ is smaller than some minimum fraction of the total number of particles N, the particle weights are degenerated and resampling should be performed.

During resampling, each particle $\hat{x}'_{t,k}$ is replaced by an existing particle $\hat{x}_{tr}$, where the index r is drawn randomly with replacement from the set $1, \ldots, N$ with probabilities ω_{tk}. The new weights are $\omega'_{t,k} = \frac{1}{N}$. It can easily be verified that the the particle distribution's moments' expectation is not changed by this operation. Furthermore, the number of effective particles N_{eff} now equals the total number of particles N.

5 Evaluating Approaches to Improving Demand Estimation

Theoretical performance guarantees can usually only be given for restricted problems and if the assumptions underlying the estimator actually fit the data. At the same time, evaluation of full real-world implementations is costly and complicated by uncertain and possibly unstable conditions. As a consequence, we measure the performance of approaches in a simulation system providing realistic but fully known and stable conditions. For this purpose, as mentioned in Section 3.1, the simulation mirrors the ARM system illustrated by Figure 1, relying on artificial demand to generate bookings. We benchmark the results using the PCRB.

5.1 Simulation ARM Model and Experimental Set-Up

As part of an ongoing research cooperation, we were given access to the simulation system REMATE as developed by Lufthansa German Airlines Gerlach et al. (2010). REMATE allows for highly flexible scenario definitions, such that real-world scenarios can be easily modeled. We adapted the system for our purposes, implementing the computation of the new estimation methods and the PCRB.

In the simulation, availabilities are optimized through dynamic programming (DP) with fare transformation. Fare transformation transforms a dependent demand model into an independent demand model so that standard independent demand optimization procedures can be applied. The dynamic programming approach explicitly models the stochastic and time-dependent nature of demand and produces a so-called bid price depending on the remaining time before departure and the remaining number of unsold seats.

The bid price is the minimum price for which the next seat is sold; all fare classes with a lower price will not be available while all fare classes with a higher price will be. See Talluri and van Ryzin (2005) for details on the DP, and Fiig et al. (2009) for a description of fare transformation.

To simplify the analysis, we restrict ourselves to a single compartment on a single flight. In the practice of airline revenue management, demand is usually forecasted separately per flight leg and compartment or per origin-destination pair and compartment. Either way, explicit consideration of multi-leg itineraries only enters the picture during optimization, which is not the focus of this paper. We therefore believe that the restriction to a single compartment and flight is not a very strong one in our context. Additionally, there are neither cancellations nor no-shows, .

Three base simulation scenarios represent domestic, continental and intercontinental markets, respectively. The number of fare classes and their prices are taken from empirical Lufthansa data for exemplary markets. Capacity is defined as 100 seats for domestic and continental flights, and as 200 seats for intercontinental flights.

For each base scenario, there exist high demand and low demand variants. In low demand variants, the capacity restriction is mostly irrelevant, such that bid prices are zero and optimization is focused on exploiting price-sensitivity. In high demand variants, bid prices are positive and optimization has to exploit price-sensitivity while being constrained by limited capacity. Each simulation includes 100 departures, to allow the estimation algorithms to settle into a stable state. Each simulation is repeated 100 times, such that the results are averages over 100 different demand realizations.

5.2 Simulation Demand Model

The simulation implements the so-called Hybrid Demand Model, where demand is decomposed into independent demand and price-dependent demand. While independent demand is defined per specific product, price-dependent demand is characterized by overall volume per itinerary, compartment, point of sale and time slice. A price elasticity parameter describes the share of customers willing to buy at a higher price.

This model has been chosen for its simplicity: With customer behavior summarized in a single elasticity parameter, result analysis and discussion are easily traceable. The addition of independent demand gives the model some flexibility to account for empirical deviations from strictly price-dependent behavior. A close variant of this model appears in the PODS revenue management simulator developed at MIT under the term Q-Forecasting Carrier (2003).

Price-sensitive demand realizes exclusively in the lowest available fare class. For that class, the expected number of bookings can be expressed as

$$x_{volume} \cdot \exp\left(-x_{elast}(\frac{f}{f_{base}} - 1)\right) \tag{28}$$

where x_{volume} and x_{elast} are the aforementioned demand parameters, f is the fare of the lowest available class and f_{base} is a reference price. Analytically, the choice of the reference price f_{base} is arbitrary, as the demand parameters x_{volume} and x_{elast} can be adapted to recover the identical choice behavior for any reference price. Setting the reference price to the lowest existing price ensures that $\exp\left(-x_{elast}(\frac{f}{f_{base}} - 1)\right) \leq 1$ such that this quantity can be interpreted as a sell-up probability. In fact, the term can be reinterpreted as a willingness-to-pay distribution, from which each customer's individual willingness-to-pay is drawn. This choice of reference price is therefore used for generating artificial customers.

Numerical stability considerations, on the other hand, suggest to use a reference price that is in the middle of the overall price range. This is done during estimation. To compare the results of the estimation methods to the actual demand parameters, the parameters have to be converted to the same reference price f_{base}. When changing the reference price f_{base} to f'_{base}, the new demand parameters are

$$x'_{volume} = x_{volume} \cdot \exp\left(-x_{elast}(\frac{f'_{base}}{f_{base}} - 1)\right) \tag{29}$$

$$x'_{elast} = x_{elast} \cdot \frac{f'_{base}}{f_{base}} \tag{30}$$

Airline industry experience shows that price elasticity will change over the booking horizon, as price-sensitive leisure customers generally book earlier than business customers, who in turn express a lower price elasticity. We therefore model the price elasticity parameter x_{elast} as a degree-two Lagrange polynomial in the square root of the number of days before departure. This function has three parameters, price elasticity at the beginning of the bookings horizon $x_{elast360}$, 60 days before departure $x_{elast60}$ and at departure

x_{elast0}.[3] At any number of days d before departure, the price elasticity is a linear combination of the three parameters.

This functional form has been chosen for having enough degrees of freedom to fit the empirical data well. The parameterization of the simulation using empirical data is described in further detail in Bartke et al. (2013).

5.3 Performance Benchmarks

To evaluate the performance of the estimation algorithms, we need benchmarks. The Posterior Cramér-Rao Bound provides the best-case scenario in terms of mean squared error, and optimization using the real demand parameters yields the optimal expected revenue. The revenue baseline is calculated through network ARM assuming strictly independent demand. To benchmark the mean squared forecast error, we introduce two additional methods, using a simple heuristic and non-linear regression, respectively, to estimate the demand parameters.

5.3.1 Posterior Cramér-Rao Bound and Information Matrix

The PCRB provides a lower bound for the mean squared error of any estimator. Formally, let $g(B_t)$ be some estimator of the demand parameters x_t operating on the booking history up to time t: $B_t = \{b_1, \dots, b_t\}$. Then, under mild regularity conditions,

$$E[(g(B_t) - x_t)(g(B_t) - x_t)^T] \geq I_t^{-1} \tag{31}$$

where "$\geq$" means that the difference between the matrices is a positive semi-definite matrix. I_t is the Fisher information matrix and evolves according to

$$I_{t+1} = M_{t+1} + (I_t^{-1} + Q)^{-1} \tag{32}$$

or, equivalently[4],

$$I_{t+1} = M_{t+1} + Q^{-1} - Q^{-1}(I_t + Q^{-1})^{-1}Q^{-1}, \tag{33}$$

[3] The choice of 360, 60 and 0 days as "anchor points" is arbitrary, however approximately equal spacing on the $\sqrt{t}$ axis improves numerical stability.

[4] To see the equivalence, note that $(A^{-1}+B)^{-1} = (1+AB)^{-1}A = B^{-1}(B^{-1}+A)^{-1}A = B^{-1}(B^{-1}+A)^{-1}(B^{-1}+A-B^{-1}) = B^{-1} - B^{-1}(B^{-1}+A)^{-1}B^{-1}$, provided that all the inverses exist.

where I_{t-1} is the Fisher information matrix from the last time step, Q is the covariance matrix from equation 1 and M_t is the Fisher information matrix of the observation at time t (Tichavsky et al. (1998) for the derivation). The measurement information M_t is defined as

$$M_t = -E[\Delta_{x_t}^{x_t} \log p_{a_t}(z_t|x_t)] \tag{34}$$

where again $\Delta_y^x = (\nabla_x)(\nabla_y)^T = (\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n})^T(\frac{\partial}{\partial y_1}, \ldots, \frac{\partial}{\partial y_n})$. Here, we assume that the real x_t are fixed, but unknown parameters. The expectation in equation 34 is therefore taken over z_t conditioned on x_t.

Using the Poisson distribution as discussed above, the likelihood $p_{a_t}(z_t|x)$ becomes[5]

$$p_{a_t}(z_t|x) = \prod_{i:h_{a_t,i}(x)>0} \frac{(h_{a_t,i}(x))^{z_t}}{z_t!} \cdot e^{-h_{a_t,i}(x)} \tag{35}$$

which yields the following measurement information matrix:

$$M_t(x) = E\left[\sum_{i:h_{a_t,i}(x)>0} \frac{(\nabla_x h_{a_t,i}(x))(\nabla_x h_{a_t,i}(x))^T}{h_{a_t,i}(x)}\right] \tag{36}$$

Equation 33 is a direct specialization of the equation given by Tichavsky et al. (1998). It efficiently computes the Fisher information matrix as it requires only one matrix inversion per iteration; Q is constant and can be inverted once at the iteration's start. For constant M_t, it is useful to also compute the steady state of the information matrix evolution, that is when $I_{t+1} = I_t$. The steady state equation

$$X = M_t + Q^{-1} - Q^{-1}(X + Q^{-1})^{-1}Q^{-1} \tag{37}$$

is a discrete, algebraic Riccati equation and there are efficient numerical methods to solve such equations (Laub (1979)).

Equation 32 is computationally less attractive, but much simpler to interpret intuitively: We compute the minimum variance of the last time step I_t^{-1}, add the additional noise introduced by the time-evolution of x_{real} with the covariance Q, convert that back to an information matrix by inversion and then add the information gained by observing at time $t+1$.

[5] We only include terms with $h_{a_t,i}(x) > 0$ in the product, that is with arrival rates > 0. If the arrival rate $h_{a_t,i}(x)$ is zero, the only possible realization is $z_{t,i} = 0$ with likelihood 1. These terms can therefore be ignored.

5.3.2 Simple Estimation

For the simple estimation approach (SE), each demand parameter is estimated independently. For each parameter, we find a value that is most consistent in least-squares sense with the current observations, holding all other parameters constant. Combining all these individual parameter estimates yields a new parameter vector which obviously "overshoots" the true parameters, since the change in each parameter alone could explain the observation. Exponential smoothing is used to avoid the overshoot. Even with exponential smoothing, non-linear choice functions can produce extreme estimates. This is avoided by using a simple outlier detection that limits the maximum change of a parameter value between time steps.

5.3.3 Maximum Likelihood Estimation

A standard solution to this type of estimation problem is maximum likelihood estimation (MLE): Given the joint probability $p_A(B, X)$, find a demand parameter trajectory $X = (x_0, \dots, x_T)$ that maximizes $p_A(B, X)$ for the observed availability and booking histories A and B. From the original state-space model in equations 2 and 3, $p_A(B, X)$ can be decomposed:

$$p_A(B, X) = p(x_0) \prod_{t=1}^{T} p_{a_t}(b_t|x_t) \cdot p(x_t|x_{t-1}) \tag{38}$$

Maximizing equation 38 (or its logarithm) over X is a very high dimensional problem. As an example, in the scenarios of this simulation study, each individual x alone has dimension 333. After 100 simulation runs, the dimension of X will be $333 \cdot 100 = 33,300$. Maximizing any non-trivial function over that many parameters is a major challenge.

To make this problem more tractable, we limit the availability and booking histories to a rolling history of a fixed number of observations (here 25). Further, we assume that x remained constant within this limited observation history and that the initial x_0 is known and equals the estimate from the preceding time step, i.e. $x_0 = \hat{x}_{t-1}$. The joint probability function then becomes

$$p_A(B, x) = p(x|\hat{x}_{t-1}) \prod_{t=1}^{T} p_{a_t}(b_t|x) \tag{39}$$

where the term $p(x|\hat{x}_{t-1})$ is given by the multivariate normal distribution with mean $\hat{x}_{t-1}$ and covariance matrix Q. The conditional probability $p_{a_t}(b_t|x)$ is the product of Poisson probability distribution functions with rates $\lambda_i = h_{a,i}(x)$. This yields the log-likelihood

$$\begin{aligned} L(x) = \log p_A(B, x) = &- \frac{1}{2}(x - \hat{x}_{t-1})^T Q^{-1}(x - \hat{x}_{t-1}) \\ &+ \sum_{\tau} \sum_{i,h_i>0} b_{\tau i} \cdot \log\left(h_{i,a_\tau}(x)\right) - h_{i,a_\tau}(x) \\ &+ const. \end{aligned} \tag{40}$$

This function is maximized by finding a root of the first derivative with the iterative Newton method. $L(x)$ is not concave in general and therefore a global maximum is not guaranteed. In practice, this does not seem to be an issue, since the maximum is expected to be close to $\hat{x}_{t-1}$ which is therefore an excellent starting value.

5.4 Forecast Initialization

In simulations, as well as in real life, forecasting methods have to be initialized in some way. As we are mainly interested in the long-term behavior of an estimation or forecasting method, the standard approach is to use a burn-in phase: the simulation is executed for a number of runs until a steady state is reached. Only after the burn-in phase are statistics collected. The initial simulations runs are therefore "wasted" computation time, and we aim to keep the required length of the burn-in phase to a minimum.

To accomplish this, we initialize the forecast with a given mean squared error, by starting with the "perfect" forecast (i.e. the real demand as known in the simulation) distorted by an error term. The solution to equation 37 from Section 5.3.1 lets us approximate the steady state PCRB I_∞^-1, which we use as the initial mean squared forecast error. This is only an approximation of the true steady state for two reasons. First, the actual mean squared forecast error will be larger than the PCRB. Second, equation 37 assumes a constant measurement information matrix $M_t = M$, which is available during actual simulation. We use a weighted average of all nested availability situations to compute an approximate M to use in equation 37.

As a result, simulations start much closer to the desired steady state. Accordingly, a short burn-in phase of 50 simulation runs seems adequate.

5.5 Numerical Results

This section presents the simulation results. Summarizing the previous sections, there are 3 demand volumes, 3 traffic types, 4 estimation methods and 10 independent demand realizations with 100 runs each, for a total of 360 simulations and 36000 simulation runs. Total running time was about 40 hours on a laptop computer with an Intel Core i5 2.6 GHz processor and 4GB of RAM.

We analyze performance in term of forecast quality and revenue. Forecast quality is measured through the so-called estimator efficiency, which is defined as the quotient of the PCRB and the mean squared error. Since there is more than one parameter to be estimated, both the PCRB and the mean squared error have the form of covariance matrices. To compute the estimator efficiency, we use the trace (the sum of diagonal elements) of these matrices. The closer the mean squared error gets to the PCRB, the higher the estimator efficiency will be. Since the PCRB is an upper bound for the mean squared error, estimator efficiency is bounded by 1 from above. Revenue results are reported as a revenue gap in percent as compared to the revenue achievable using the real demand parameters as a “perfect” forecast.

5.5.1 Forecast Quality

Figures 2, 3 and 4 show the estimator efficiency obtained in the simulation. For each data point, the estimated mean and approximate 95% confidence intervals are given.[6] Figure 2 provides an aggregated overview over all demand volumes and scenarios. SE produces a mean squared error that is orders of magnitude higher than the PCRB and thus efficiency is close to 0. The mean squared error of PF is about 3 times as high as the PCRB, which leads to an estimator efficiency that is lower than that of the UKF and MLE. The UKF is more efficient than MLE, but this result is just below the threshold to statistical significance.

Figure 3 splits the data by demand volume. All methods suffer from a loss of efficiency in the high demand case. That loss is least pronounced for PF, but its mean squared error fluctuates strongly over independent simulation runs, leading to the large size of the confidence interval.

[6] Here the data points are in fact ratios of two experiments. We use the R-package `pairwiseCI` (Froemke et al. (2012)), which implements Ogawa (1983), to compute the approximate confidence intervals.

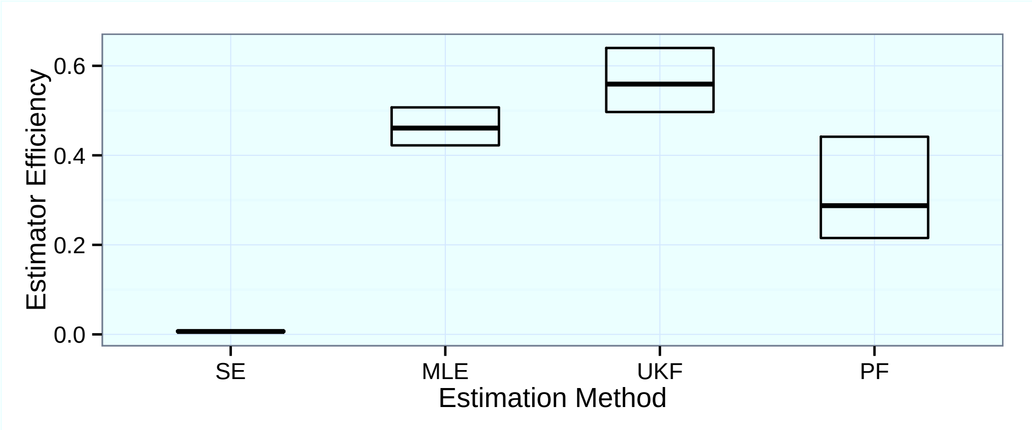

Figure 2: Estimator efficiency $E = {}^{tr(\text{PCRB})}/_{tr(\text{MSE})}$, aggregated over all scenarios; the boxes represent the mean value and its 95% confidence interval; SE = Simple Estimation, MLE = Maximum Likelihood Estimation, UKF = Unscented Kalman Filter, PF = Particle Filter

A possible explanation is quality of availability information degrading as the demand volume increases. If bid prices are zero, availability is solely determined by fare transformation. These availabilities will thus be stable, providing perfect availability information to the estimator. If bid-prices are positive, availability may change each time a booking occurs and every time the bid price vector gets updated (once per day in the simulation). The estimators, however, are not aware of these availability changes, they only get a rough approximation[7] of the total amount of time a class was available during a time slice.

Figure 4 splits the data by traffic type. For intercontinental traffic all methods show reduced estimator efficiency compared to the other two scenarios, an effect that is statistically significant for MLE and UKF, but not for PF. Qualitative results are the same for all three scenarios, only MLE has a slightly, but not significantly, higher estimator efficiency than UKF in the “Domestic” scenario.

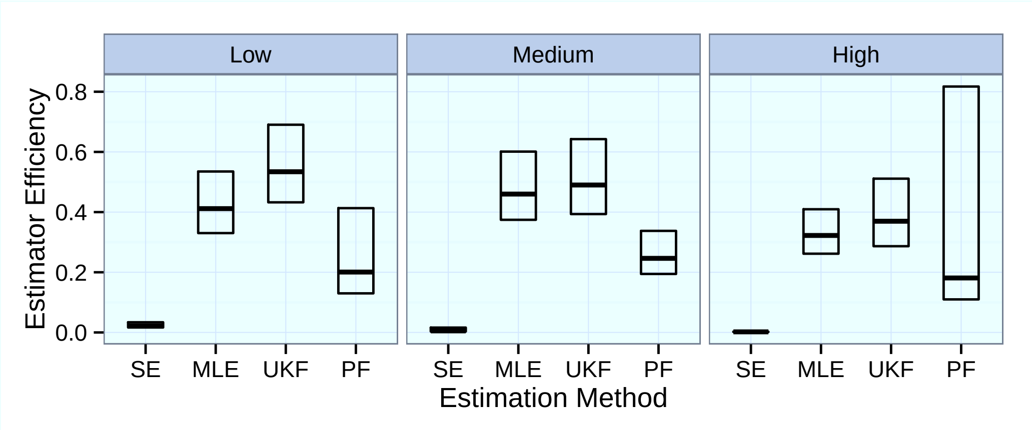

Figure 3: Estimator efficiency $E = {}^{tr(\text{PCRB})}/_{tr(\text{MSE})}$ by demand volume; the boxes represent the mean value and its 95% confidence interval; SE = Simple Estimation, MLE = Maximum Likelihood Estimation, UKF = Unscented Kalman Filter, PF = Particle Filter

5.5.2 Revenue

Figures 5 through 7 show the relative gap in revenue of each estimation method compared to the revenue obtained given a "perfect" forecast. Figure 5 displays the aggregated data over all traffic types and demand volumes, while figures 6 and 7 provide more detail by splitting the data by demand volume and traffic type, respectively. Again, for each data point the mean and its 95% confidence interval are shown.

SE yields significantly less revenue than all other methods. In the aggregate over all scenarios, MLE achieves the highest revenues, with only a slight gap ($\approx 0.25\%$) compared to the upper bound. UKF and PF are close to the MLE, however the gaps are comparatively almost twice ($\approx 0.4\%$) and ten times ($\approx 2.2\%$) as high. Considering the 95% confidence intervals, the distinction between the UKF and the MLE is not statistically significant.

The more detailed views in figures 6 and 7 confirm that the qualitative results are the same over all demand volumes and traffic types. This suggests that our results and conclusions are robust under various perturbations of the scenario parameters.

[7] based on a linear interpolation of the bid prices at the beginning and the end of a time slice

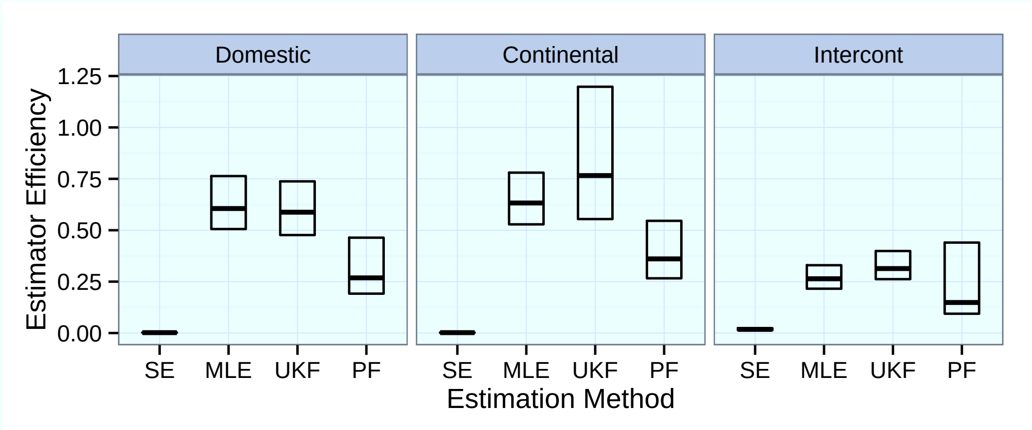

Figure 4: Estimator efficiency $E = {}^{tr(\text{PCRB})}/_{tr(\text{MSE})}$ by traffic type; the boxes represent the mean value and its 95% confidence interval; SE = Simple Estimation, MLE = Maximum Likelihood Estimation, UKF = Unscented Kalman Filter, PF = Particle Filter

6 Conclusion

Based on a control-theory view of airline revenue management, this paper formulated the ARM process as a state-space model. We employed this model to adapt and apply two Filter techniques, Unscented Kalman Filter and Particle Filter, to create new demand estimates. In the previous section, these estimates were evaluated against the PCRB and alternative estimates. In this final section, we further discuss the results and provide an outlook to future research in this direction.

6.1 Discussion of Results

The results presented in Section 5 unambiguously show that SE is clearly inferior to its alternatives, both in terms of forecast error and revenue performance. From a practitioner's standpoint, the large revenue gap is especially troublesome. Note, however, that the size of this gap is partially due to the scenario setup emphasizing the role of price-sensitivity.

MLE, on the other hand, has a slight advantage in both metrics over all other methods. We suspect that this is due to the relatively slow demand evolution found in the data used to calibrate the scenarios. Since MLE par-

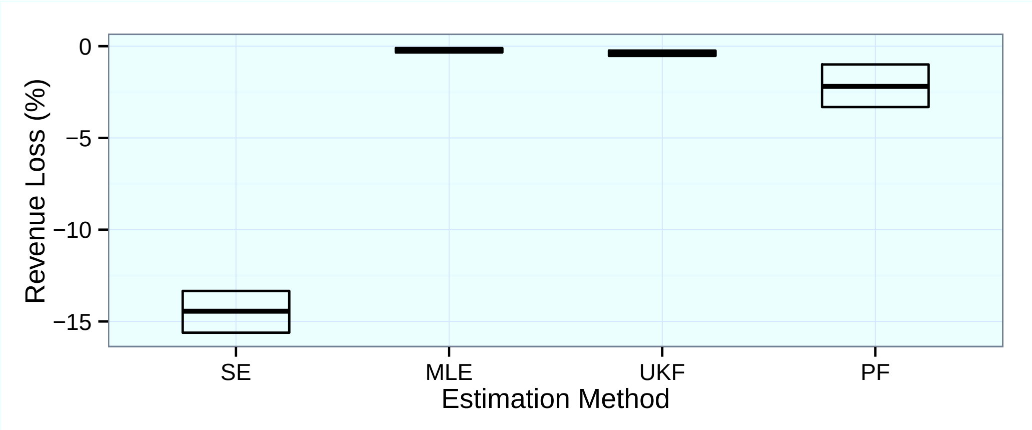

Figure 5: Revenue gap in percent compared to revenue given a perfect forecast, aggregated over all scenarios; the boxes represent the mean value and its 95% confidence interval; SE = Simple Estimation, MLE = Maximum Likelihood Estimation, UKF = Unscented Kalman Filter, PF = Particle Filter

tially assumes constant demand, its performance should degrade compared to the Filter estimates when demand changes faster.

Among the Filter estimates, UKF performs notably better than PF. Increasing the number of particles used might change this result. However, with the given settings, PF computation time already exceeded that of UKF by a factor of about two. In real-world implementations, computation time would certainly be a concern, so significantly increasing the number of particles does not appear as a realistic option.

Vulcano et al. (2009)) estimate that the revenue gain from using choice-based revenue management and MLE is between 1% and 5%. In our simulation study, we can see that using a simple, ad-hoc estimation method can easily negate that potential, especially when the airline moves towards a restriction-less fare structure. Carefully designed estimation methods on the other hand can perform close enough to the optimum that a positive revenue effect from choice-based revenue management still exists.

6.2 Research Outlook

Our particular demand model is quite restrictive and was mostly chosen for illustrative purposes. However, we believe that our approach of modeling

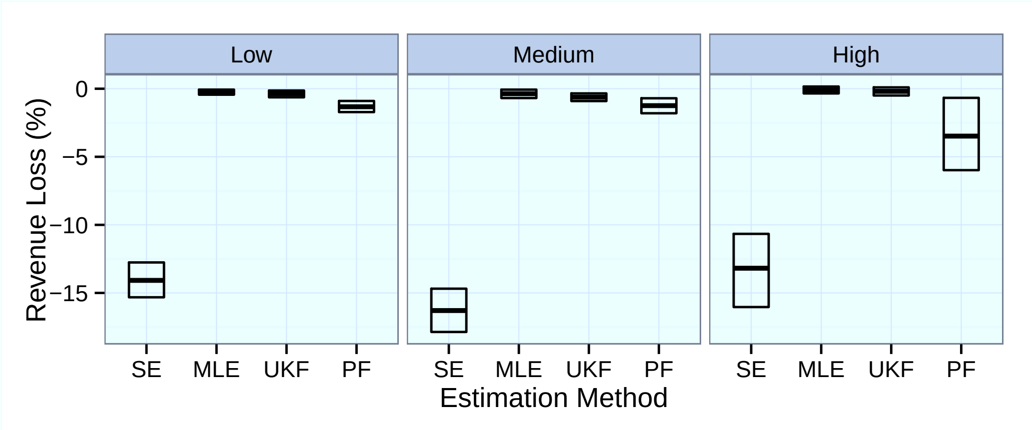

Figure 6: Revenue loss in percent compared to revenue under a perfect forecast, by demand volume; the boxes represent the mean value and its 95% confidence interval; SE = Simple Estimation, MLE = Maximum Likelihood Estimation, UKF = Unscented Kalman Filter, PF = Particle Filter

the demand estimation problem in a state-space framework offers a path to quickly develop high quality estimation methods for any given demand model. There exists a large body of literature on many variants of the state-space estimation problem, in many cases with a strong focus on practical applicability. Once the state-space formulation exists, it is a matter of searching the literature for suitable methods and adapting them accordingly. Therefore, other extensions of our model, e.g. away from the Poisson distribution or the simple AR(1) model, seem also within reach.

Making use of the additional information provided by the covariance of the estimates is another interesting topic for future research. Here, we focus our attention on using covariance information to improve forecast quality when the total number of sales from which demand has to be estimated is small. In real-world implementations, it is a natural choice to consider origin-destination pairs as distinct markets for which separate demand parameters are estimated. However, in a typical airline's network, the distribution of sales over these origin-destination pairs is highly non-uniform, with a small number of these markets concentrating most of the sales. In turn, for the majority of markets only a very small number of bookings will be observed in a given time-period, rendering high-quality demand estimation at this level impossible. The availability of covariance information helps to identify

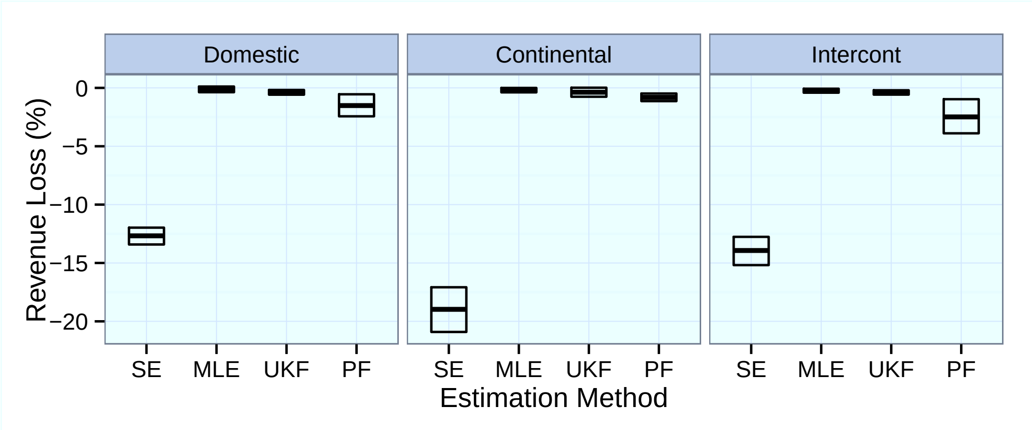

Figure 7: Revenue loss in percent compared to revenue under a perfect forecast, by traffic type; the boxes represent the mean value and its 95% confidence interval; SE = Simple Estimation, MLE = Maximum Likelihood Estimation, UKF = Unscented Kalman Filter, PF = Particle Filter

those problematic markets and we are currently investigating a clustering procedure in which demand estimates from multiple markets are merged using a variance-weighted mean.

Acknowledgements

We thank the anonymous referees for their helpful comments as well as Lufthansa German Airlines for access to the simulation system REMATE and empirical calibration data.

References

Aviv, Y., Pazgal, A., 2005. Dynamic Pricing of short life-cycle products through active learning, http://en.scientificcommons.org/42458091.

Bartke, P., Cleophas, C., Kliewer, N., 2013. Working Paper: Data-Driven Parameterization of Simulation Systems for Airline Revenue Management, submitted to BISE Special Issue Decision Analytics.

Belobaba, P. P., 1989. Application of a Probabilistic Decision Model to Airline Seat Inventory Control. Operations Research 37 (2), 183–197.

Bitran, G., Wadhwa, H., 1996. A methodology for demand learning with an application to the optimal pricing of seasonal products.

Carrier, E., 2003. Modeling Airline Passenger Choice: Passenger Preference for Schedule in the Passenger Origin-Destination Simulator (PODS). Master's thesis, Massachusetts Institute of Technology.

Carvalho, A., Puterman, M., 2004. How should a manager set prices when the demand function is unknown.

Chung, B., Li, J., Yao, T., Kwon, C., Friesz, T. L., 2012. Demand Learning and Dynamic Pricing under Competition in a State-Space Framework. Engineering Management, IEEE Transactions on 59 (2), 240–249.

Doucet, A., Godsill, S., Andrieu, C., 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. Statistics and Computing 10 (3), 197–208.

Fiig, T., Isler, K., Hopperstad, C., Belobaba, P., 2009. Optimization of mixed fare structures: Theory and applications. Journal of Revenue and Pricing Management 9 (1), 152–170.

Froemke, C., Hothorn, L., Schneider, M., 2012. pairedCI: Confidence intervals for the ratio of locations and for the ratio of scales of two paired samples.

Gerlach, M., Cleophas, C., Frank, M., 2010. Introducing REMATE: Revenue management simulation in practice. In: Meeting of the AGIFORS Working Group Revenue Management and Cargo, New York.

Gordon, N., Salmond, D., Smith, A., 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. Radar and Signal Processing, IEE Proceedings F 140, 107–113.

Horner, P., 2000. Mother, father of invention produce golden child: Revenue management. OR/MS Today, 47.

Julier, S. J., Uhlmann, J. K., 1997. A New Extension of the Kalman Filter to Nonlinear Systems. In: Signal processing, sensor fusion, and target recognition VI; Proceedings of the Conference. pp. 182–193.

Kalman, R. E., 1960. A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME-Journal of Basic Engineering 82 (Series D), 35–45.

Kitagawa, G., 1996. Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. Journal of Computational and Graphical Statistics 5 (1), 1–25.

Kwon, C., Friesz, T. L., Mookherjee, R., Yao, T., Feng, B., 2009. Noncooperative competition among revenue maximizing service providers with demand learning. European Journal of Operational Research 197 (3), 981–996.

Laub, A., 1979. A Schur method for solving algebraic Riccati equations. IEEE Transactions on Automatic Control 24 (6), 913–921.

Li, J., Yao, T., Gao, H., 2009. A Revenue Maximizing Strategy Based on Bayesian Analysis of Demand Dynamics. In: SIAM Proceedings: Mathematics for Industry. Society for Industrial and Applied Mathematics, San Francisco, CA, pp. 174–181.

Littlewood, K., 1972. Forecasting and control of passenger bookings. In: AGIFORS 12th Annul Symposium Proceedings October, Nathanya, Israel. pp. 193–204.

Lobo, M. S., Boyd, S., 2003. Pricing and learning with uncertain demand. Preprint, Fuqua School of Business, Duke University.

Müller, P., 1991. Monte Carlo Integration in General Dynamic Models. In: Statistical Multiple Integration: Proceedings of an AMS-IMS-SIAM Joint Research Conference. pp. 145–162.

Ogawa, J., 1983. On the confidence bounds of the ratio of the means of a bivariate normal distribution. Annals of the Institute of Statistical Mathematics 35 (1), 41–48.

Sen, A., Zhang, A. X., 2009. Style goods pricing with demand learning. European Journal of Operational Research 196 (3), 1058–1075.

Stefanescu, C., 2009. Multivariate Customer Demand: Modeling and Estimation from Censored Sales. SSRN Electronic Journal.

Talluri, K. T., van Ryzin, G., 2005. The theory and practice of revenue management. International series in operations research & management science. Springer.

Tichavsky, P., Muravchik, C. H., Nehorai, A., 1998. Posterior Cramér-Rao Bounds for Discrete-Time Nonlinear Filtering. IEEE Transactions on Signal Processing 46 (5), 1386–1396.

Vulcano, G., van Ryzin, G., Chaar, W., 2009. OM Practice – Choice-Based Revenue Management: An Empirical Study of Estimation and Optimization. Manufacturing & Service Operations Management 12 (3), 371–392.

Vulcano, G., van Ryzin, G., Ratliff, R., 2012. Estimating Primary Demand for Substitutable Products from Sales Transaction Data. Operations Research 60 (2), 313–334.