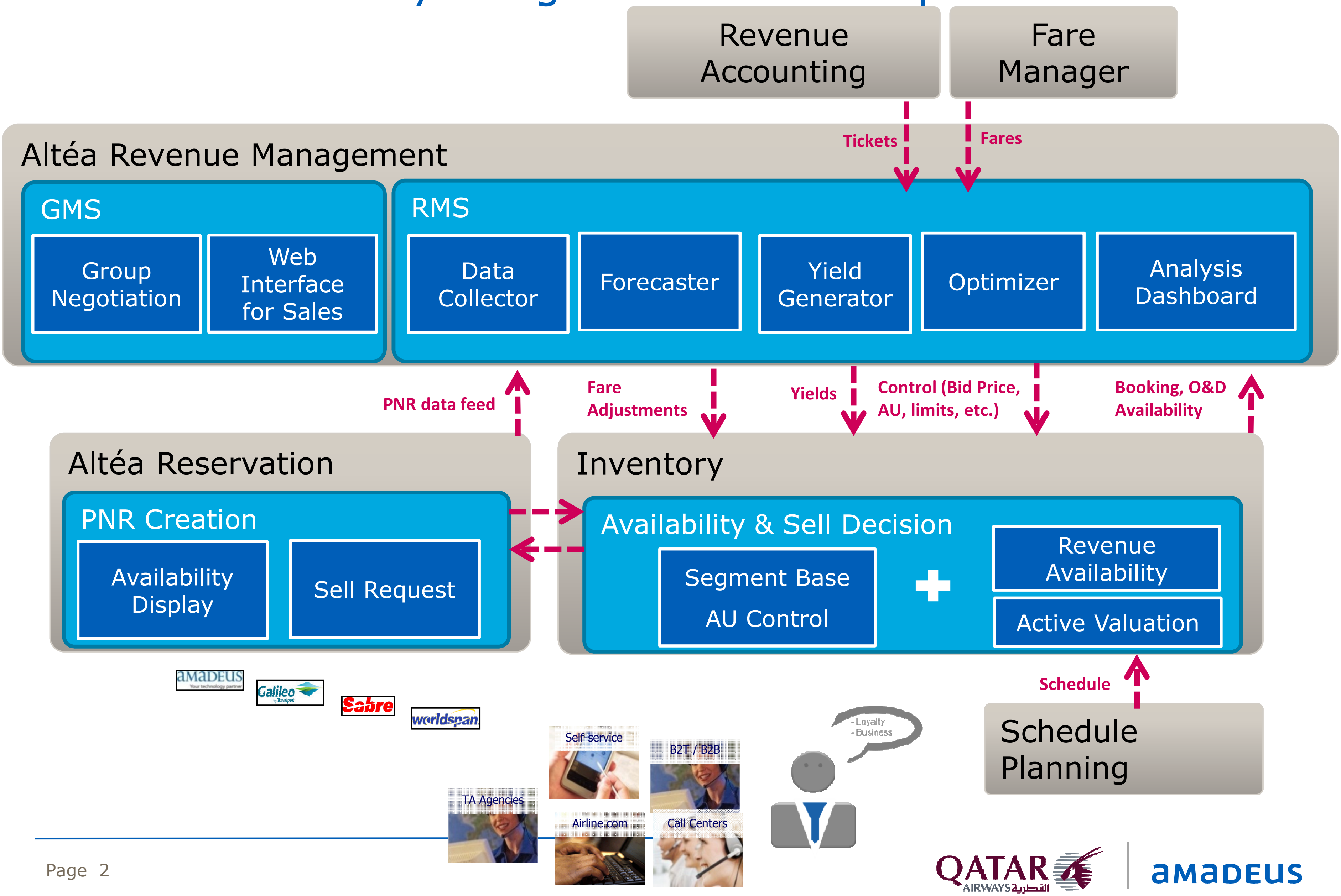
Scientific Corner - Forecasting methodology

Thomas Fiig

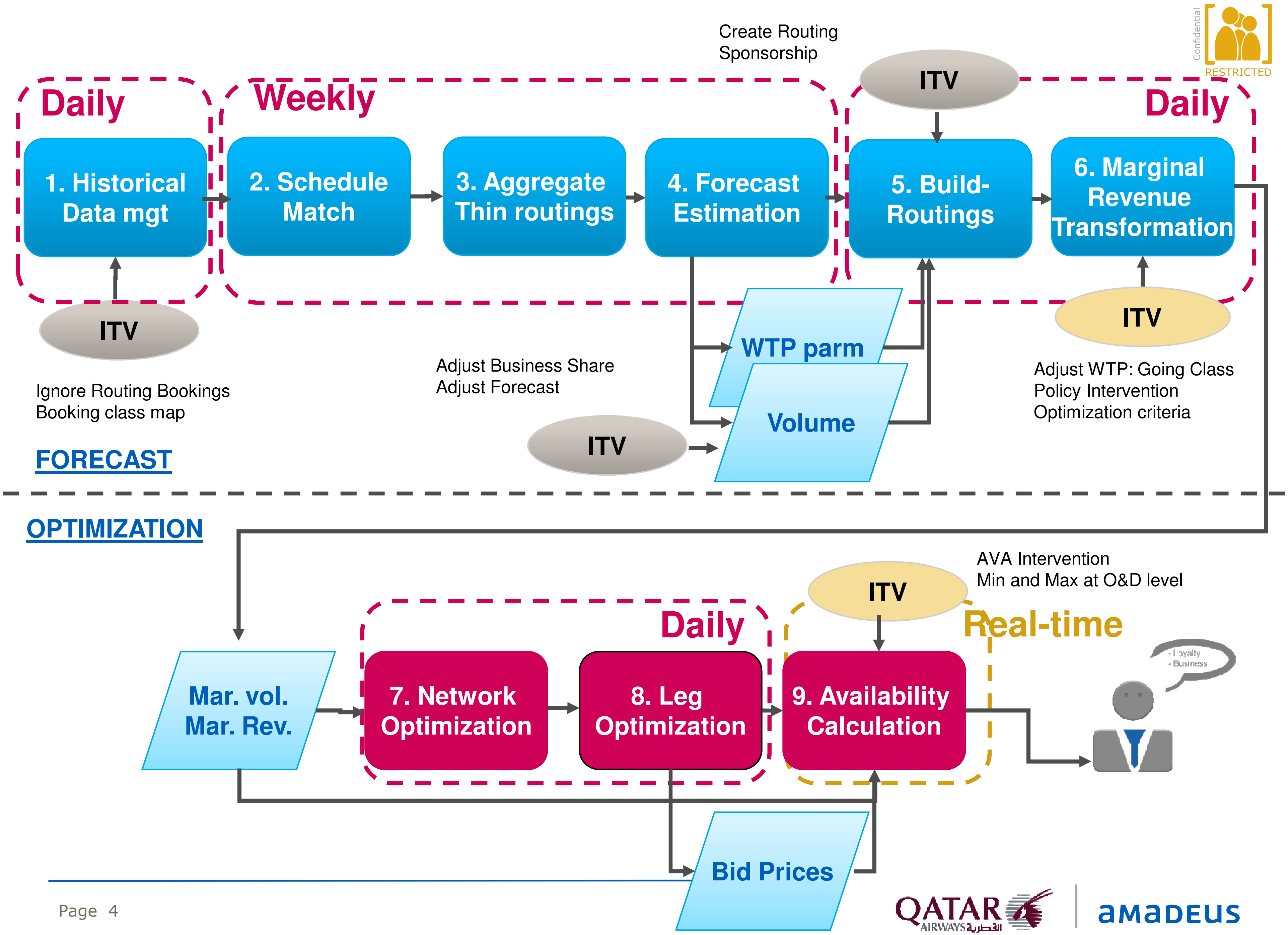
Doha, Oct 3-4 2016

RM Solutions fully integrated into Altéa platform



Confidential
RESTRICTED

1 Forecasting - Overview



Create Routing
Sponsorship
Daily
Weekly
Daily
ITV
1. Historical Data mgt
2. Schedule Match
3. Aggregate Thin routings
4. Forecast Estimation
5. Build-Routings
6. Marginal Revenue Transformation
ITV
ITV
WTP parm
Volume
ITV
Ignore Routing Bookings
Booking class map
Adjust Business Share
Adjust Forecast
Adjust WTP: Going Class
Policy Intervention
Optimization criteria
FORECAST
OPTIMIZATION
AVA Intervention
Min and Max at O&D level
ITV
Daily
Real-time
Mar. vol.
Mar. Rev.
7. Network Optimization
8. Leg Optimization
9. Availability Calculation
Bid Prices

Definition of the traffic flow

Confidential
RESTRICTED

Schedule

Legs	OSLCPH CPHSFO
Departure time	13:25 15:35
Point of Sale	DK
Point of Origin	DK
Direction	Outbound
Travel purpose	Leisure
Passenger type	I
Previous airlines	DY
Next airlines	UA
Previous cities	BGO
Next cities	LAX

Routing

Historical Data Management

Confidential
RESTRICTED

Input Data

Reservations
- Historic PNR

Inventory (traffic flow level)
- Interventions (Segment limits)
- Current Inventory counters
- Availability per depdate, dbd.
- Schedule information

DCS
- No-show

ETS
- Ancillary revenue

Revenue Accounting
- Coupons (Tickets information)

Fare Quote
- Fares (ATPCO), Tax, Airline specific tax

Airline Data
- Carrying Cost (variable cost)

Market Data
- Competitor prices (QL2/Infare)
- Total Demand
- Schedules

Schedule data are shared instantaneously to all RM modules.

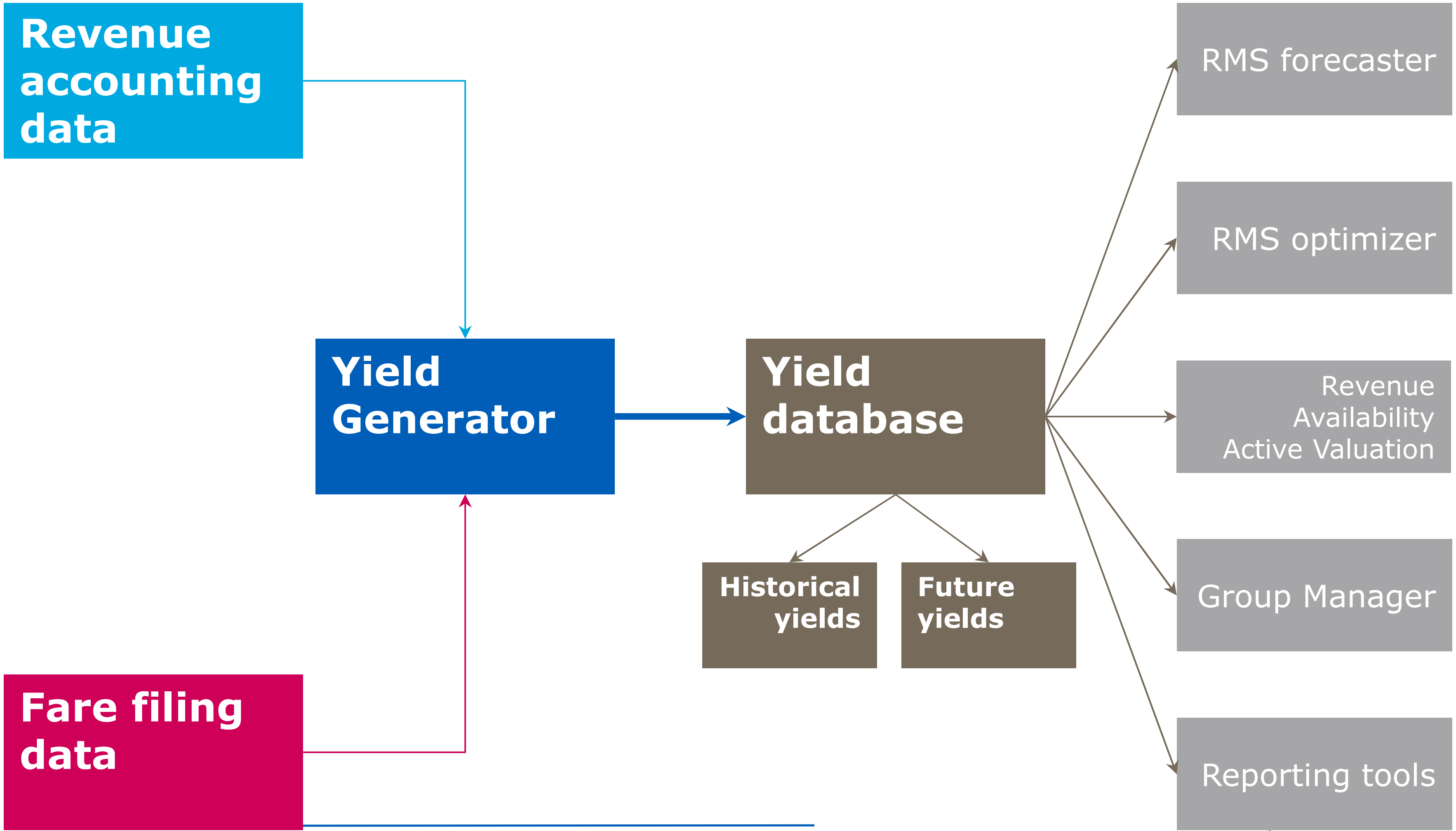
Taxes and fare rules interpretation requires unique knowledge of pricing.

Confidential
RESTRICTED

2 Yield estimation

Yield Estimation: High level view

Yield generation overview



Confidential RESTRICTED

Data sources

Yield generation overview

Revenue accounting data

- Historical coupon information with historical price and proration records
- Objectives:
  - Compute **historical yields** from historical prices
  - Size passenger traffic to weight fares in the computation of **future yields**

Fare Filing data

- Prices currently on sale (fares+surcharges+taxes), retrieved by Priced Fare Matrix
- Objectives:
  - Estimate what the customer will pay
  - Derive **future yields** from new prices on sale

Confidential
RESTRICTED

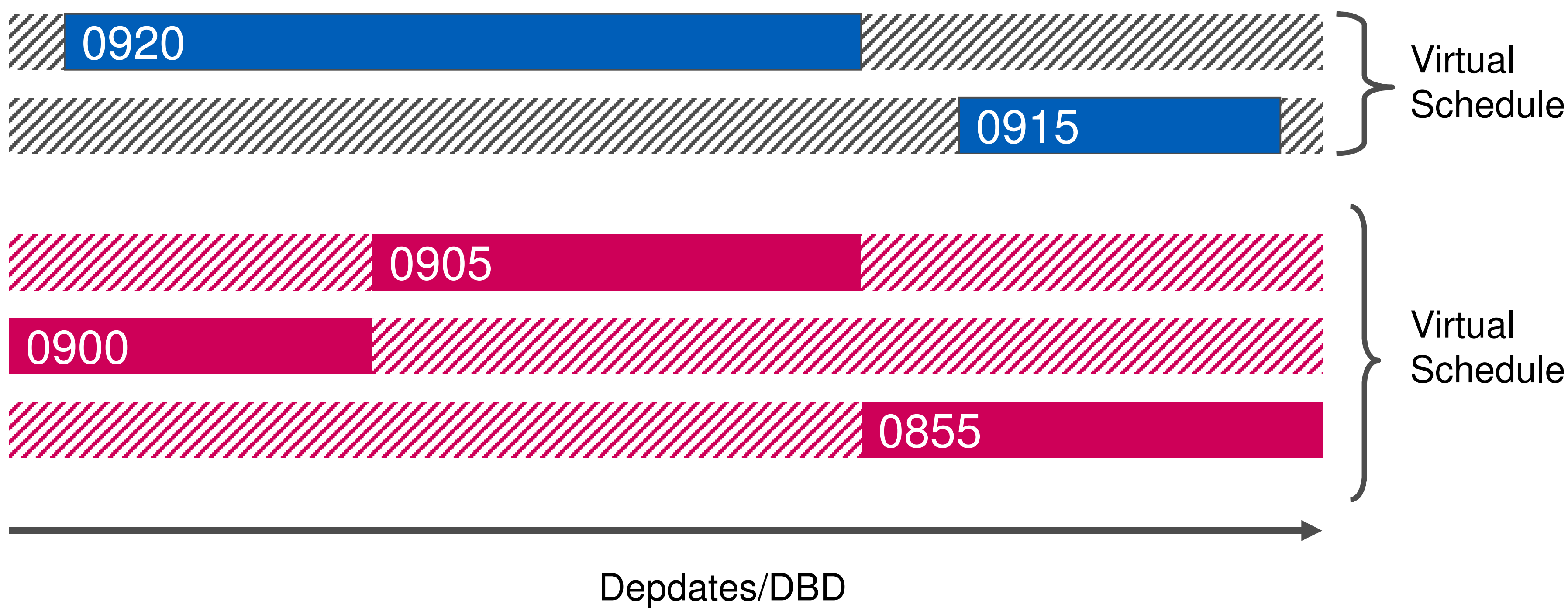
3 Forecasting – Schedule Match

Historical Data Management

Determine Significant Historical Flows

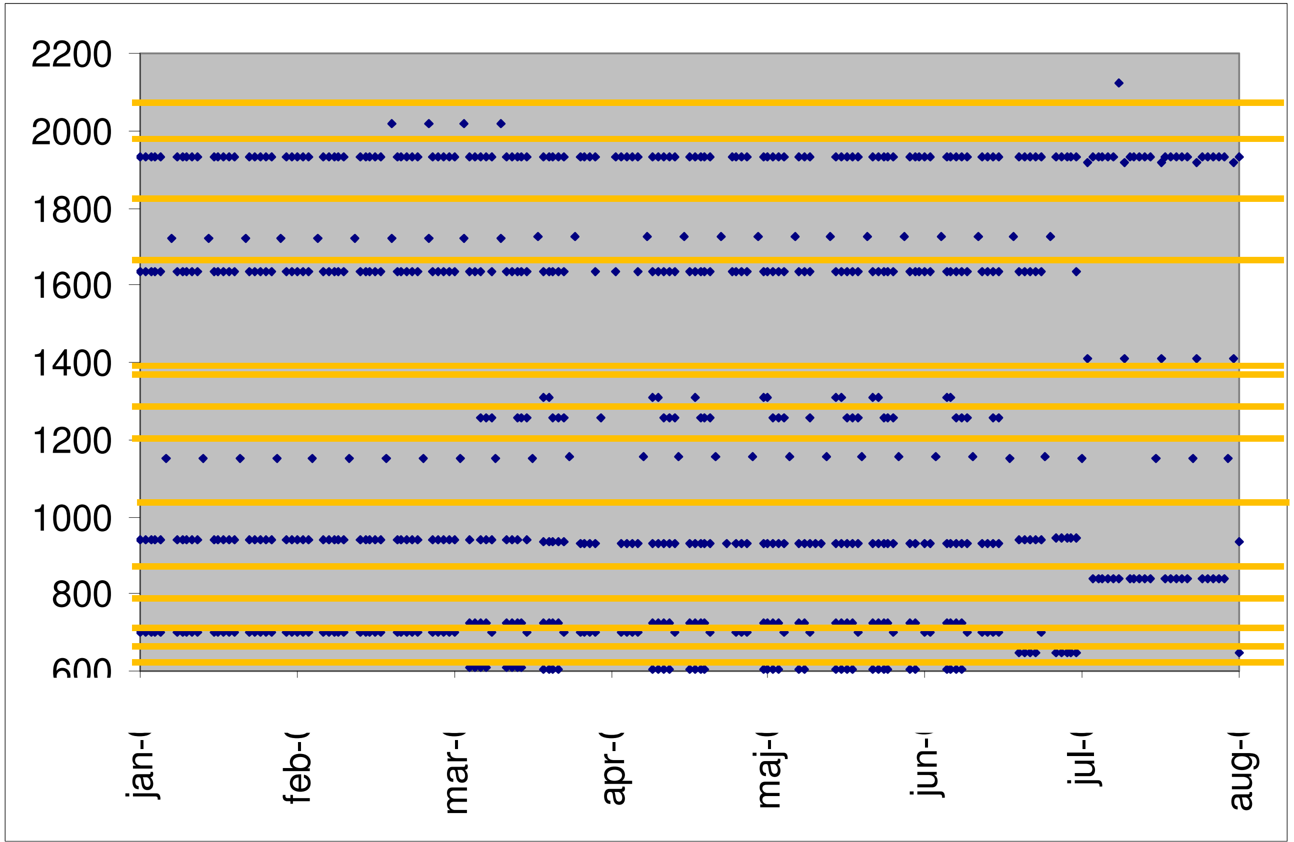
- Virtual schedule is an sequence of flight schedules. (VSID)
- Schedules within a virtual schedule cannot be overlapping.
- Each routing is uniquely mapped to a virtual schedule.

Actual flight schedules



Schedule Match Algorithm

Single leg



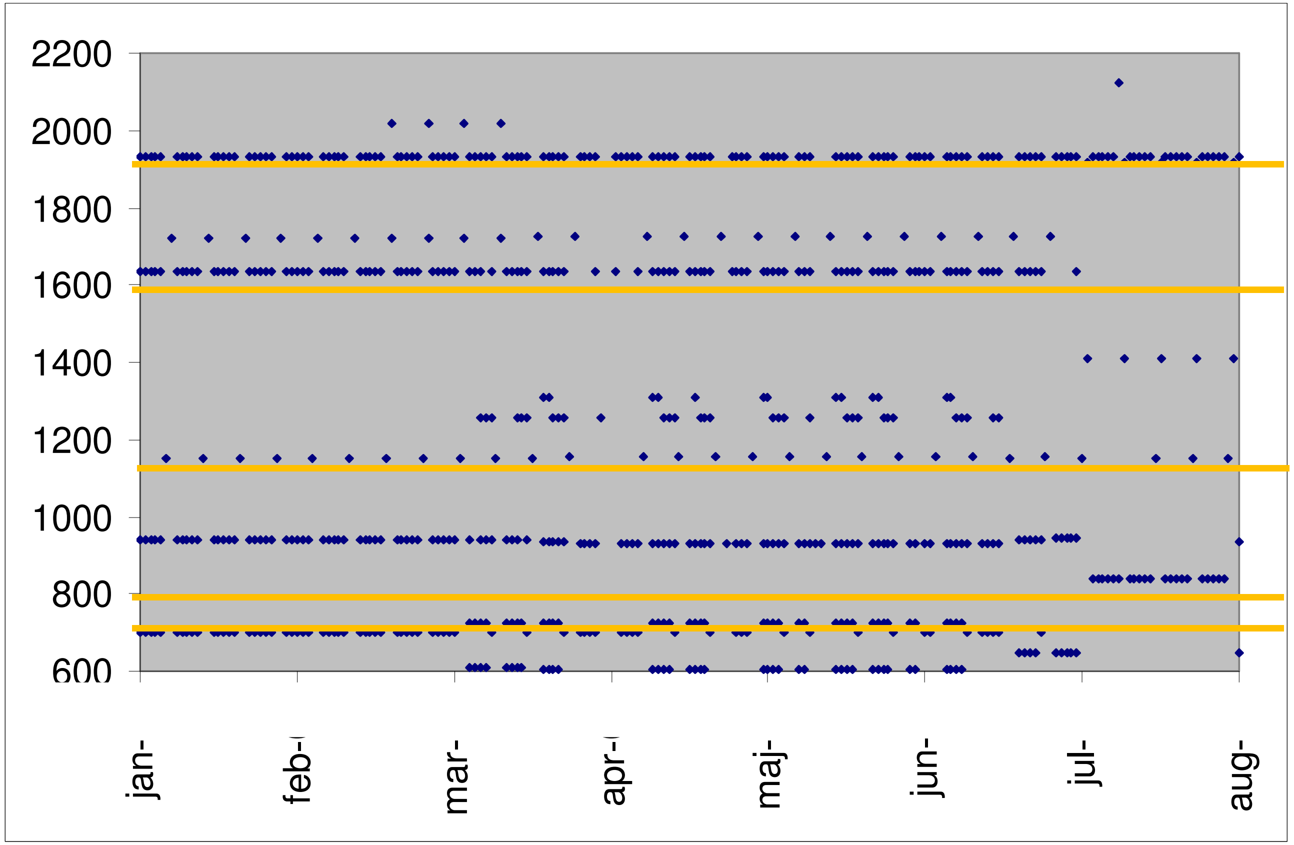
- Schedules are joined into virtual schedules.

Algorithm:

- Rank schedules according to sum of pax.
- Determine boundary as largest range where there are no overlap.
- Join schedules and repeat.

Schedule Match Algorithm

Single leg

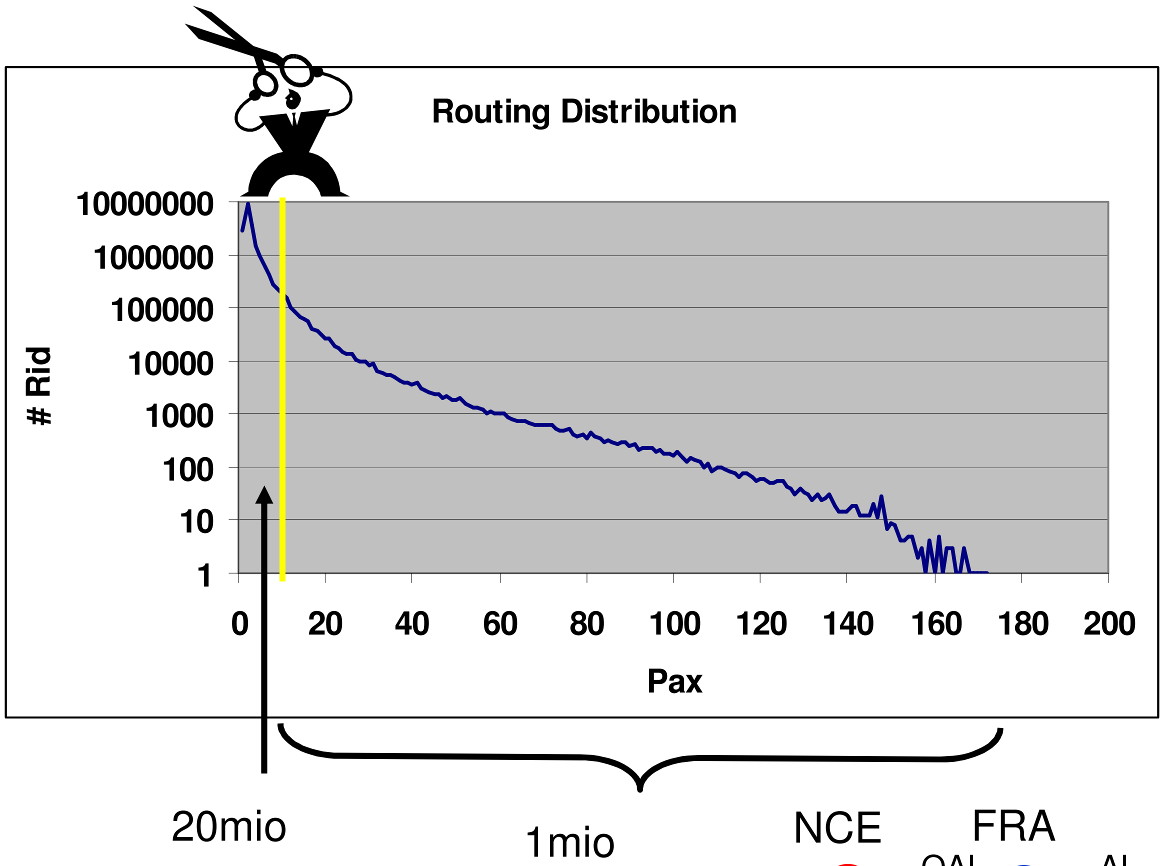


- Schedules are joined into virtual schedules.

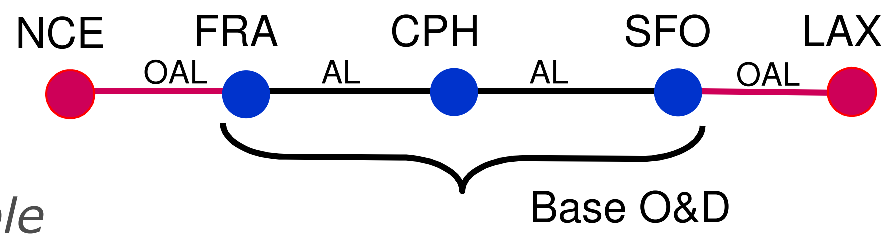
Algorithm:

- Rank schedules according to sum of pax.
- Determine boundary as largest range where there are no overlap.
- Join schedules and repeat.

Aggregating thin routings to base O&D



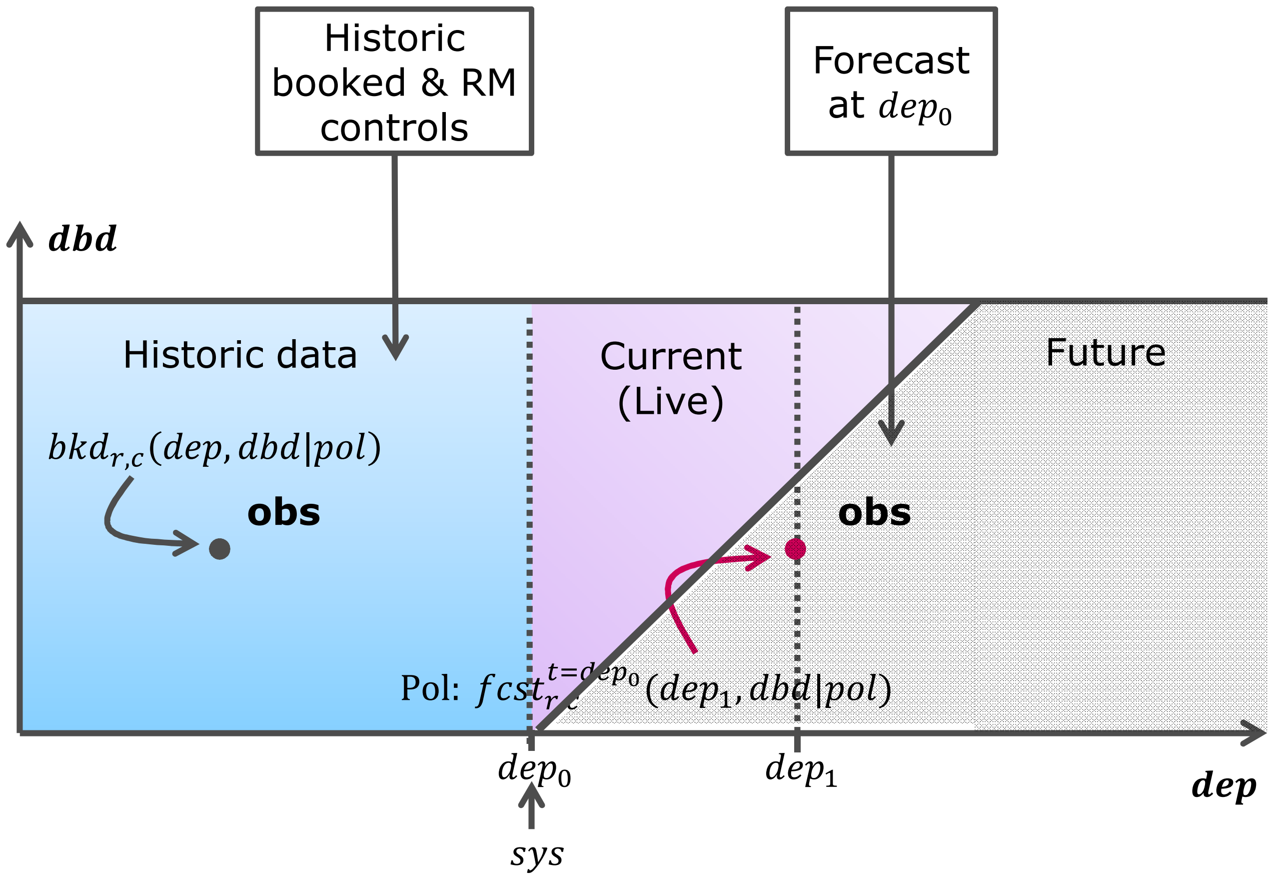
- Thin/Thick based on less/more than 5 pax per two years history.
- Cut thin routings to base O&D
- Remove OAL Prev/Next airline and city
- Point of Sale = “YY”
- POC = Control O&D On city



Note: Number of flows is scalable to airline size

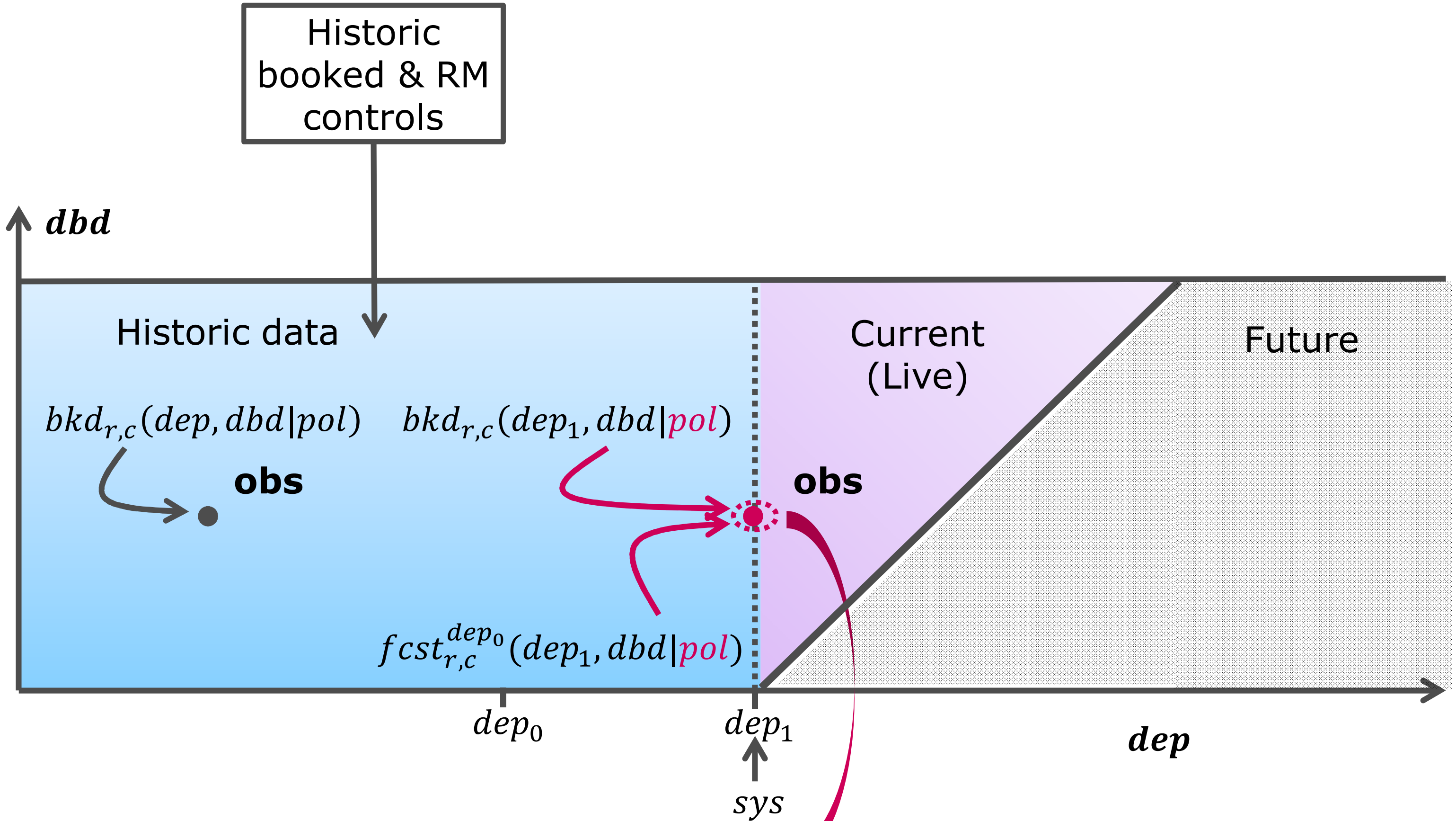
4 Forecasting – Estimation

Forecast Quality Measurements



Confidential RESTRICTED

Forecast Quality Measurements



$$err_{r,c}^{dep_0} = fcst_{r,c}^{dep_0}(dep_1, dbd|pol) - bkd_{r,c}(dep_1, dbd|pol)$$

Parametric Forecasting Model

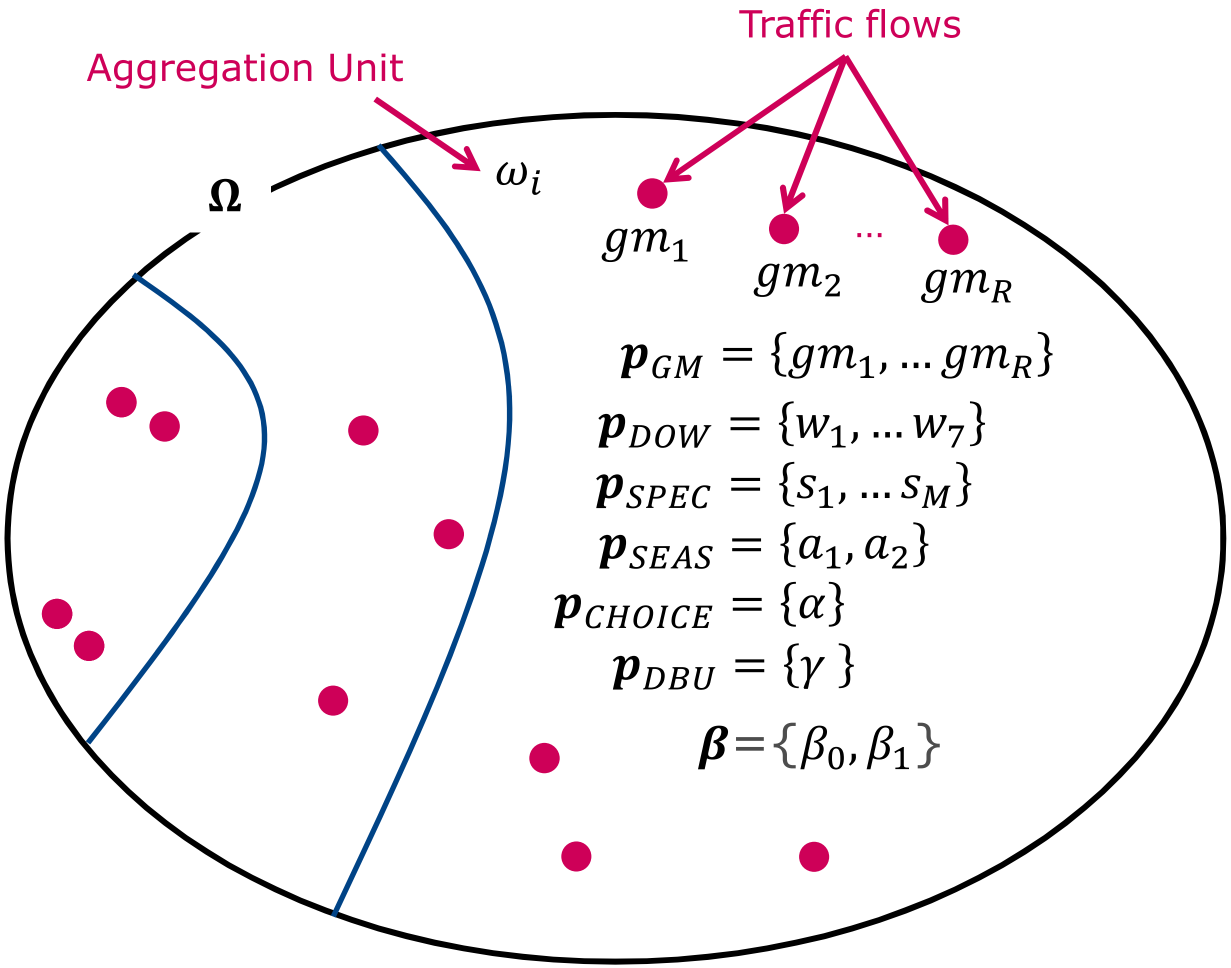
Single fare family

$$fcst_{r,c}(\boldsymbol{p}, dep, dbd \mid pol = \{c\}) = vol(\boldsymbol{p}, dep, dbd) \times MNL(\boldsymbol{f}, \boldsymbol{\beta}) =$$

$$\underbrace{\overset{(1)}{gm_r}\ \overset{(2)}{w_k}\,\overset{(3)}{s_m}\overset{(4)}{\left(1 + a_1 \sin\left(2\pi \frac{doy}{365}\right) + a_2 \cos\left(2\pi \frac{doy}{365}\right)\right)}\ \overset{(5)}{e^{-\gamma \cdot \frac{dbd}{dbd_c}}(1 - e^{-\gamma})}}_{\text{Volume}}\ \underbrace{\overset{(6)}{MNL(\boldsymbol{f}, \boldsymbol{\beta})}}_{\text{Choice prob.}}$$

	Description	Factor	Parameters	# Par.
(1)	**Grand mean**	$f_{GM,r}(\boldsymbol{p}_{GM}) = gm_r;\ r = 1, \dots, R$	$\boldsymbol{p}_{GM} = (gm_1, \dots, gm_R)$	R
(2)	**Day of week**	$f_{DOW}(\boldsymbol{p}_{DOW}, dep) = w_k;\ k = 1, \dots, 7$	$\boldsymbol{p}_{DOW} = (w_1, \dots, w_7)$	6
(3)	**Special periods**	$f_{SPEC}(\boldsymbol{p}_{SPEC}, dep) = s_m;\ m = 1, \dots, M$	$\boldsymbol{p}_{SPEC} = (s_1, \dots, s_M)$	$M - 1$
(4)	**Seasonality**	$f_{SEAS}(\boldsymbol{p}_{SEAS}, dep) = \left[1 + a_1 \sin\left(2\pi \frac{doy}{365}\right) + a_2 \cos\left(2\pi \frac{doy}{365}\right)\right]$	$\boldsymbol{p}_{SEAS} = (a_1, a_2)$	2
(5)	**Differential build-up**	$f_{DBU}(\boldsymbol{p}_{DBU}, dbd) = e^{-\gamma \cdot dbd}(1 - e^{-\gamma})$	$\boldsymbol{p}_{DBU} = \gamma$	1
(6)	**Choice prob. (1 fare fam.)**	$f_{PROB,c}(\boldsymbol{p}_{PROB} \mid pol = \{c'\}) = \mathrm{e}^{V(f,\boldsymbol{\beta})}/(1 + \mathrm{e}^{V(f,\boldsymbol{\beta})})$	$\boldsymbol{\beta} = (\beta_{FF}, \beta_f)$	2

Parameters levels



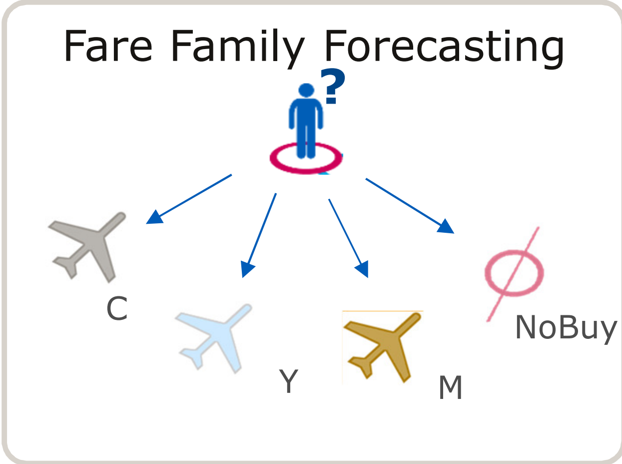
- Traffic flows partitioned into aggregation units.
- Within agg. unit only one volume parameter per flow. All other parameters shared by the agg. unit.
- Number of free vol. parameters: R+M+9.
- Number of free beta parameters:
  1 preference per FF.,
  1 elasticity per passenger segment.
- **Robust**: Parsimonious use of parameters, provides robustness for sparse data.

Choice probability

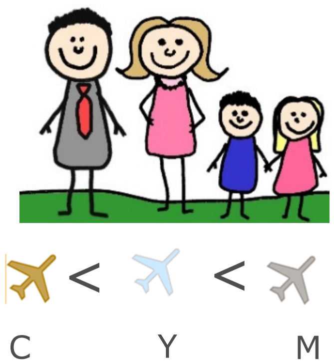
MNL model in Altéa RMS by customer type

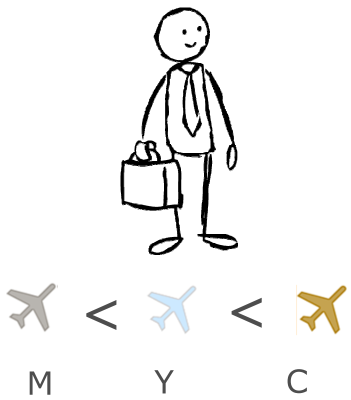
A potential customer can choose between different alternatives: buy product (in Fare Family CYM) or NoBuy.

$$p_C + p_Y + p_M + p_{NoBuy} = 1$$



Choice determined by the perceived relative utility of alternatives. The utility is different for everyone.
Depends partly on individual specific factors such as if travelling for work, or whether bringing children on the trip.





$$V_i = \beta_{0,i} - \beta_1 \cdot f_i$$

$$p_i = \frac{e^{V_i}}{\sum_j e^{V_j}}$$

Estimation

Estimation methodology log likelihood

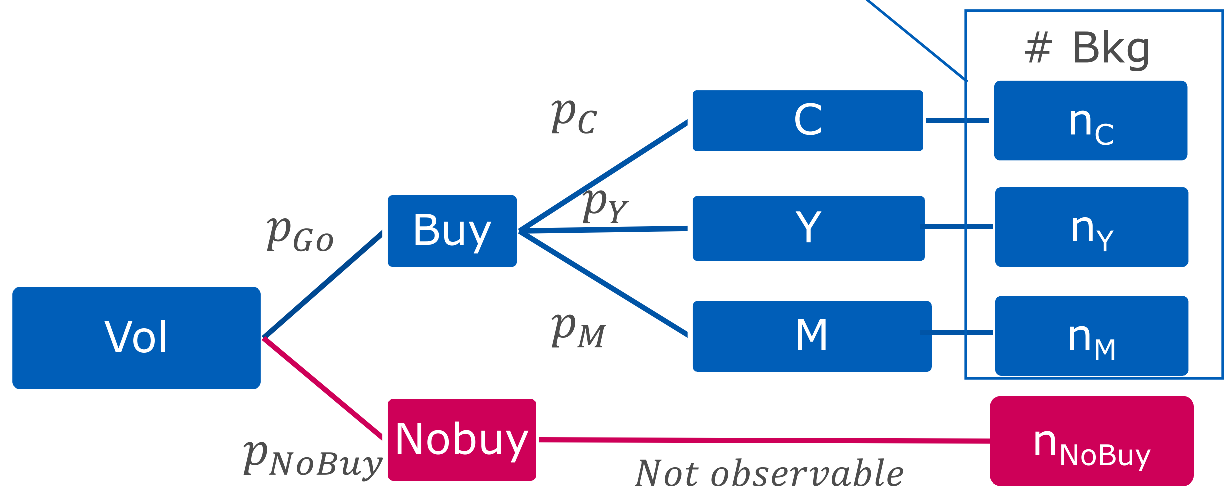
- Assume the arrival rate is Poisson distributed; and the choice prob. follow a multinomial distribution => Total likelihood of model parameters is the product of the two distributions, given the historical booking observations in C,Y, and M.

$$Likelihood(Vol,\boldsymbol{\beta} \mid n_C, n_Y, n_M, n_{NoBuy}) = Pois(Vol)\ \mathrm{Multinom}(\boldsymbol{\beta}, n_C, n_Y, n_M, n_{NoBuy})$$

- It can be simplified into a log-likelihood function:

$$LL = \sum_{obs}\left(n_{tot} \ln(fcst) - fcst + \sum_{i \in C,Y,M} n_i \ln(p_i^*)\right)$$

p_i^* is the prob of buying the alternative i *given* that the customer buys something*)



*) Jeffrey P. Newman, Mark E. Ferguson, Laurie A. Garrow, Timothy L. Jacobs (2014) Manufacturing & Service Operations Management 16(2):184-197.

Solution $(\partial LL_\omega/\partial \boldsymbol{p})(\boldsymbol{p}^*) = \boldsymbol{0}, (\partial LL_\omega/\partial \boldsymbol{\beta})(\boldsymbol{\beta}^*) = \boldsymbol{0}$

Solve week index k=1,..,7 $\quad \frac{\partial LL}{\partial w_k} = 0$

Solve special index m=1,..,M $\quad \frac{\partial LL}{\partial s_m} = 0$

Solve WTP β $\quad \frac{\partial LL}{\partial \beta} = 0$

Solve season a_1, a_2 $\quad \frac{\partial LL}{\partial a_1} = 0, \frac{\partial LL}{\partial a_2} = 0$

Solve grand mean r=1,..,R $\quad \frac{\partial LL}{\partial gm_{vrid}} = 0$

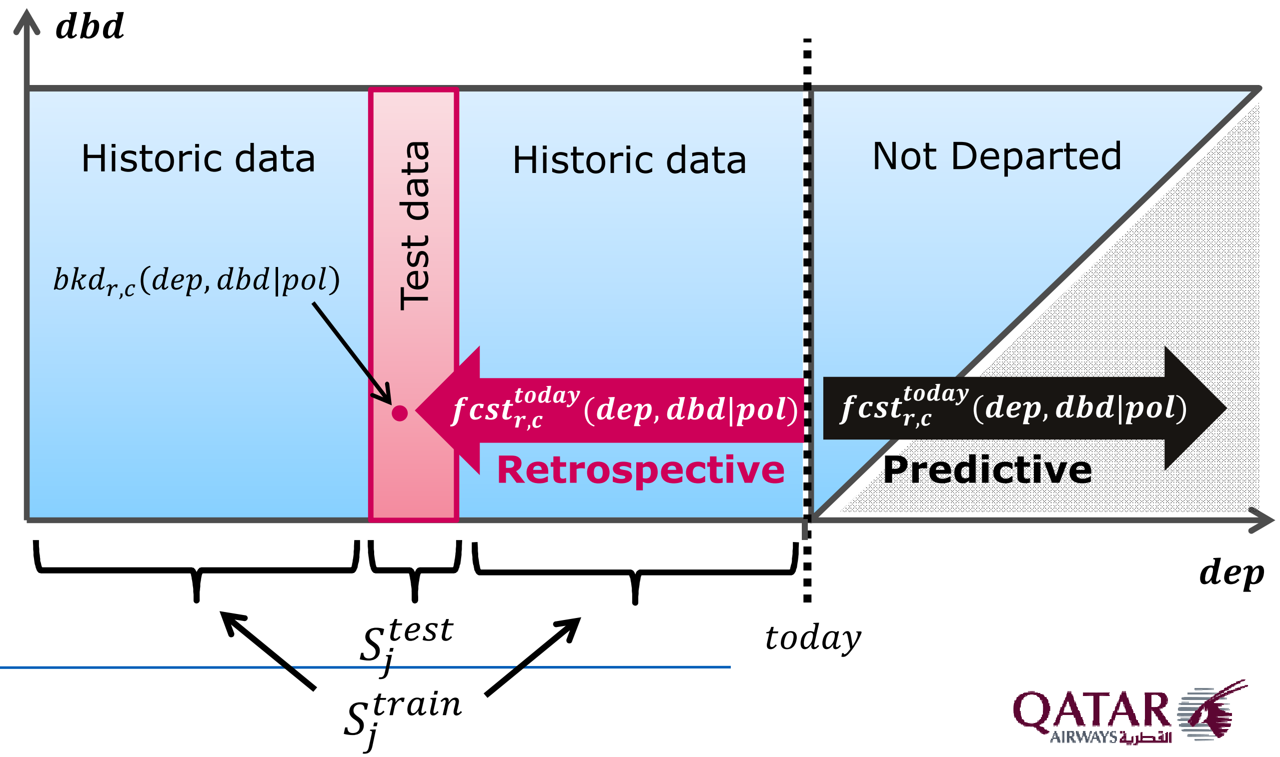
Iterate until convergence

- **Consistent solution**: All forecast components solved simultaneously.
- **Separable equations**: Due to parametric model are decomposed into separate components.
- **Unique optimum**: Each forecast component is convex in the unknown parameters.
- **Analytical solution**: Properties of MSE and BIAS error measures enable an analytical solution.
- **Advantages**:
  - Robustness
  - Accuracy
  - Computational Speed

Real Airline Data

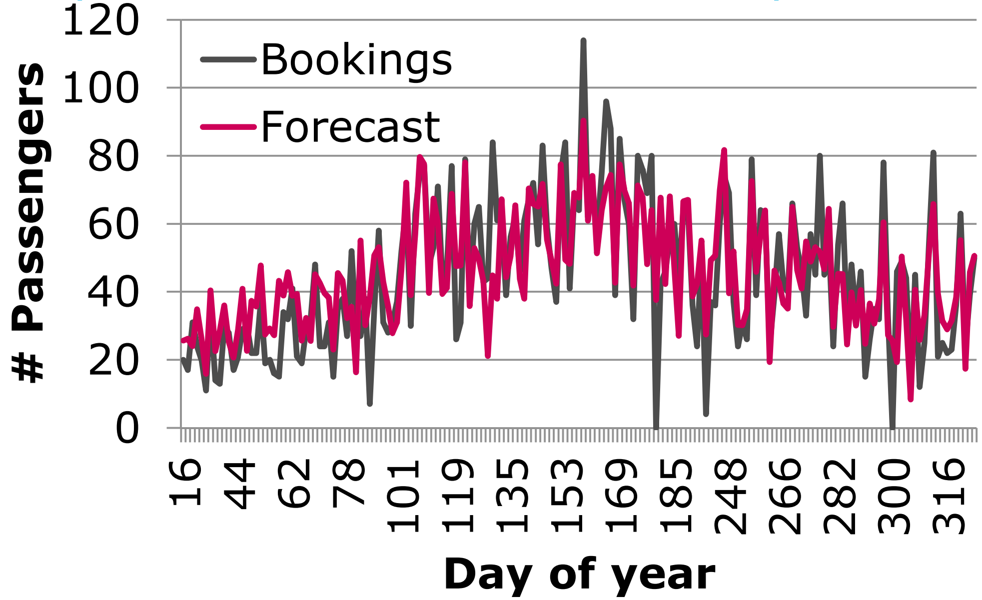
Jack-knife resampling techniques on retrospective forecast

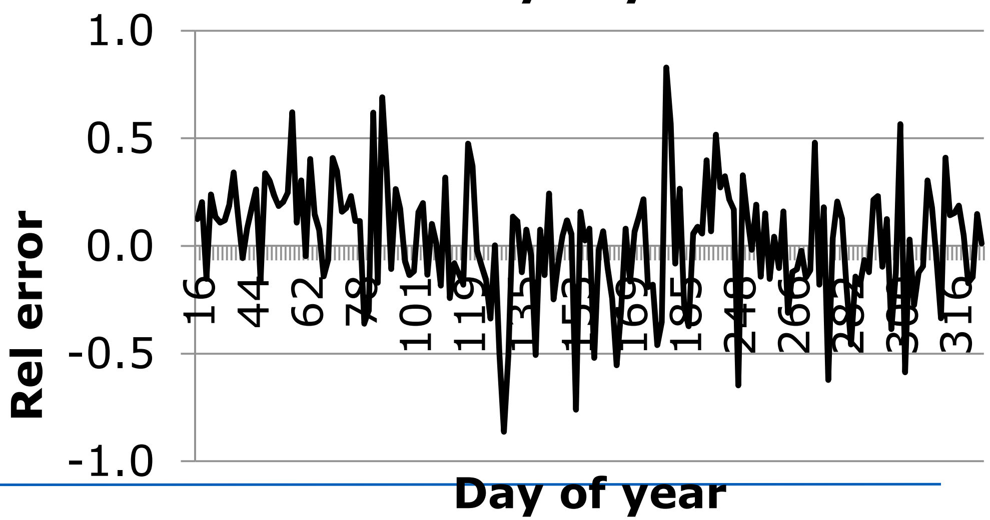
- Jack-knife resampling techniques
  - Construct N resampled datasets from the original sample data, by repeatedly splitting into a training S_j^{train} and a test dataset S_j^{test} based on the departure dates.
- Compare retrospective forecast with historical observations
  - Compare at the lowest level. Aggregate in various dimensions.



Real Airline Data

Comparison of booked and retrospective forecast

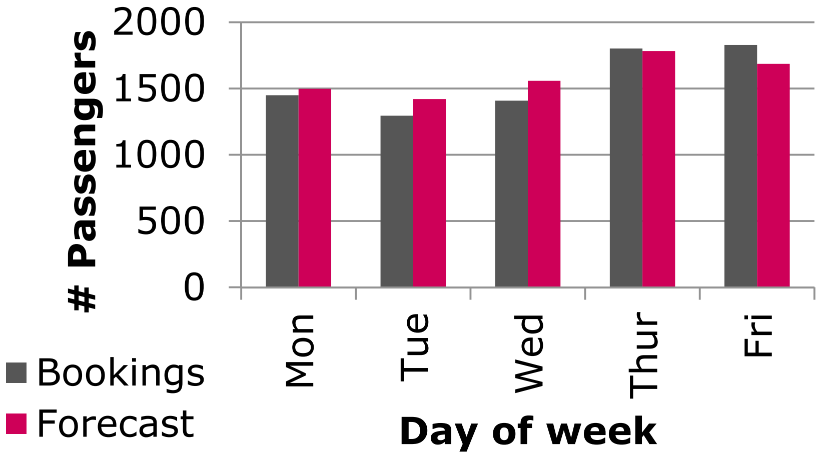


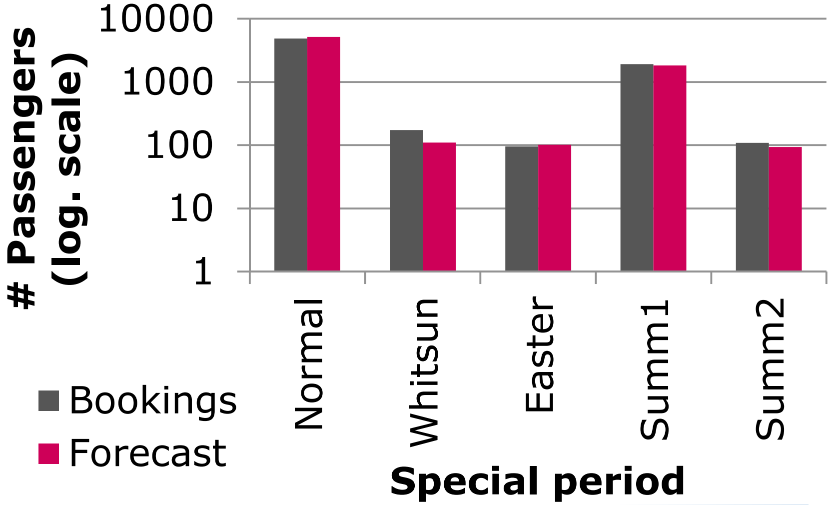


- Total no. of book: 7,780.
- Total no. of obs.: 1.4×10^6
- Total no. parameters: 33
- **No systematic errors**: The forecasting model accurately captures the underlying factors.

Real Airline Data

Comparison of booked and retrospective forecast (continued)

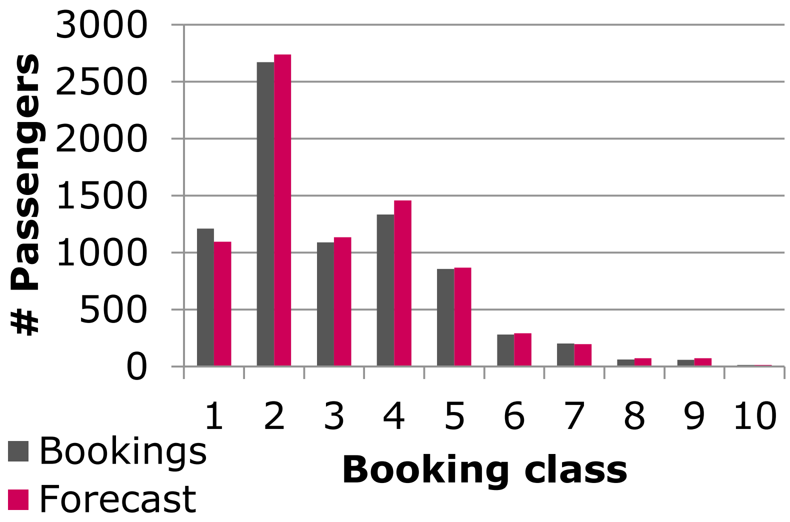


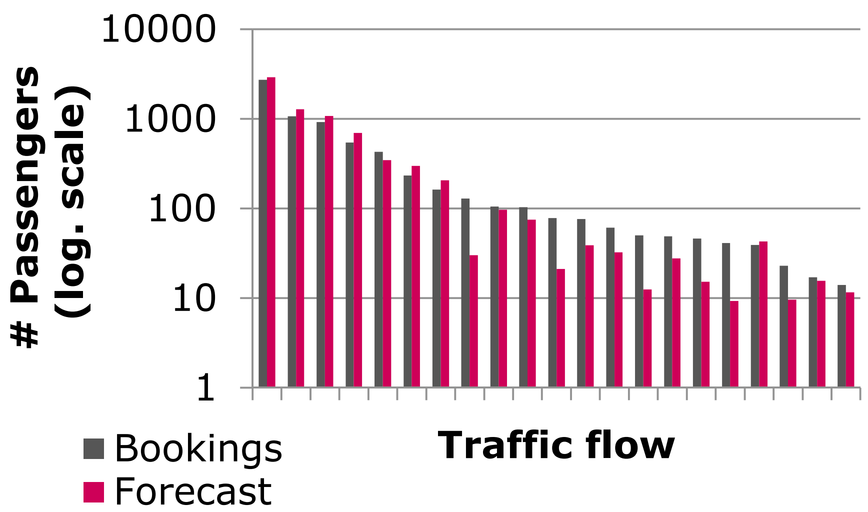


- **Excellent forecast quality for all forecast components**: Observed here for DOW and Special period.

Real Airline Data

Comparison of booked and retrospective forecast (continued)





- **Excellent forecast quality across class dimension:** Provides accurate estimates of the sell-up forecast parameters
- **Excellent forecast quality for flow volume**: Observed here over two decades.

5 Build Routings

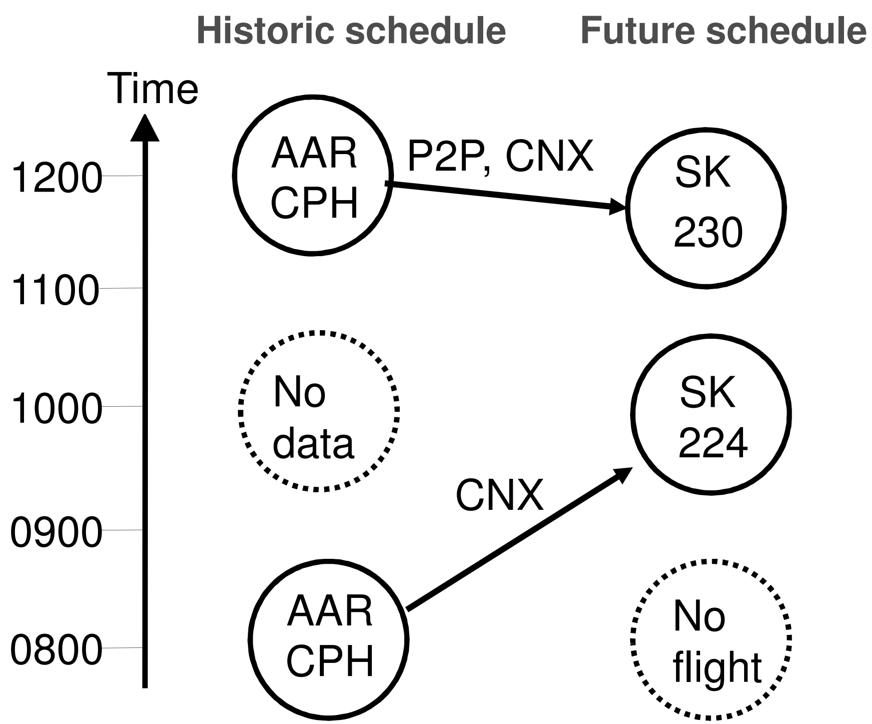
Build Routings

Rules based

5. Build-Routings

Matching the future schedules with historical schedules.

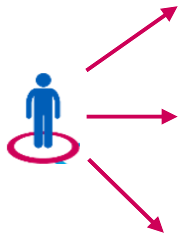
- Builds (all) feasible O&D schedules incl. OAL schedules and selects closest match to historic departed.
- New flights (DOW) automatically handled by forecast parameters.
- New destinations/frequencies: Automatic sponsors based on proximity.
- Dead-ends. Re-route dead-ends to alternative if appropriate.
- Live PNRs used to create new traffic flows.
- Interventions used to create user defined CNX



Research - Improved Build Routings

Confidential
RESTRICTED

- Different travel solutions possible.
- A potential customer can choose between different alternatives.
- Demand is split between the alternatives.
- Choice determined by the perceived relative utility of alternatives.



$p_1 = 80\%$
AMSNRT

$p_2 = 18\%$
AMSTPE
TPENRT

$p_3 = 2\%$
AMSBKK BKKTPE
TPENRT

Trip Duration (T)
All flights and Layovers

Number of Stops (S)
Or number of connections

Trigonometric Time Preferences
(depTime, arrTime)

→ Schedule Utility V

$$p_i = \frac{e^{V_i}}{\sum_j e^{V_j}}$$

$$V_i = \beta_s \cdot S_i + \beta_t \cdot T_i + \sum_{k=1}^{3} \beta_{sk} \cdot \sin(2k\pi t) \quad + \beta_{ck} \cdot \cos(2k\pi t)$$

amadeus

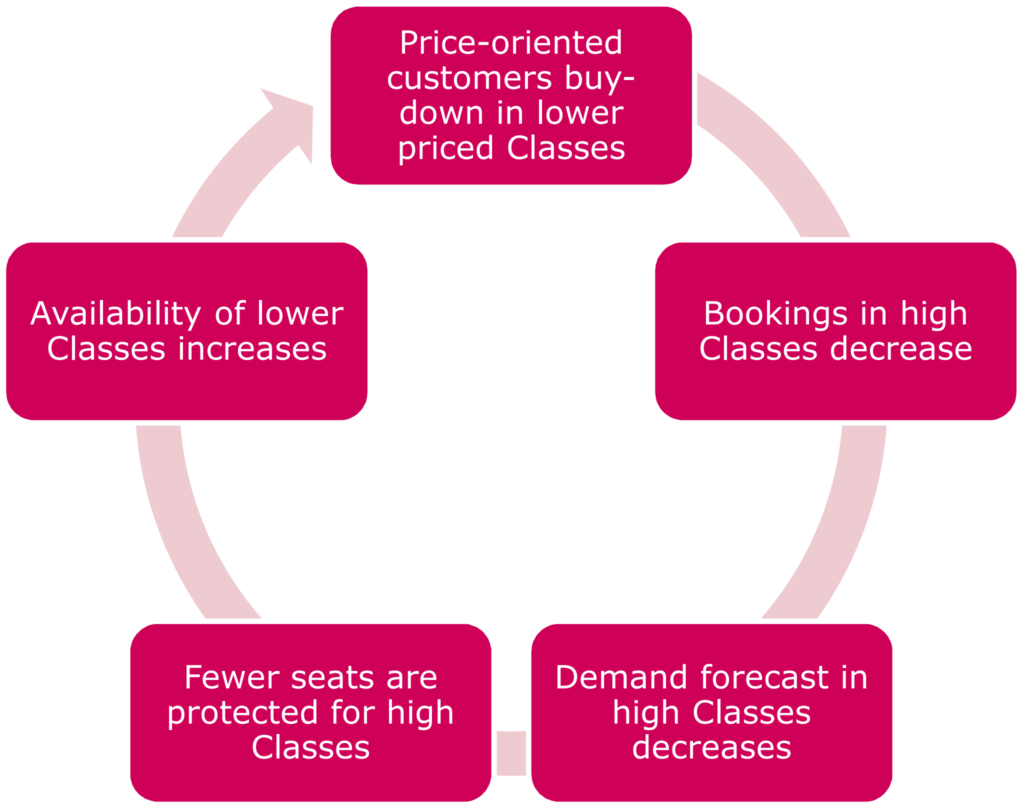
6

Marginal Revenue Transformation

Simplified fare-structures → Spiralling Down

6. Marginal Revenue Transformation

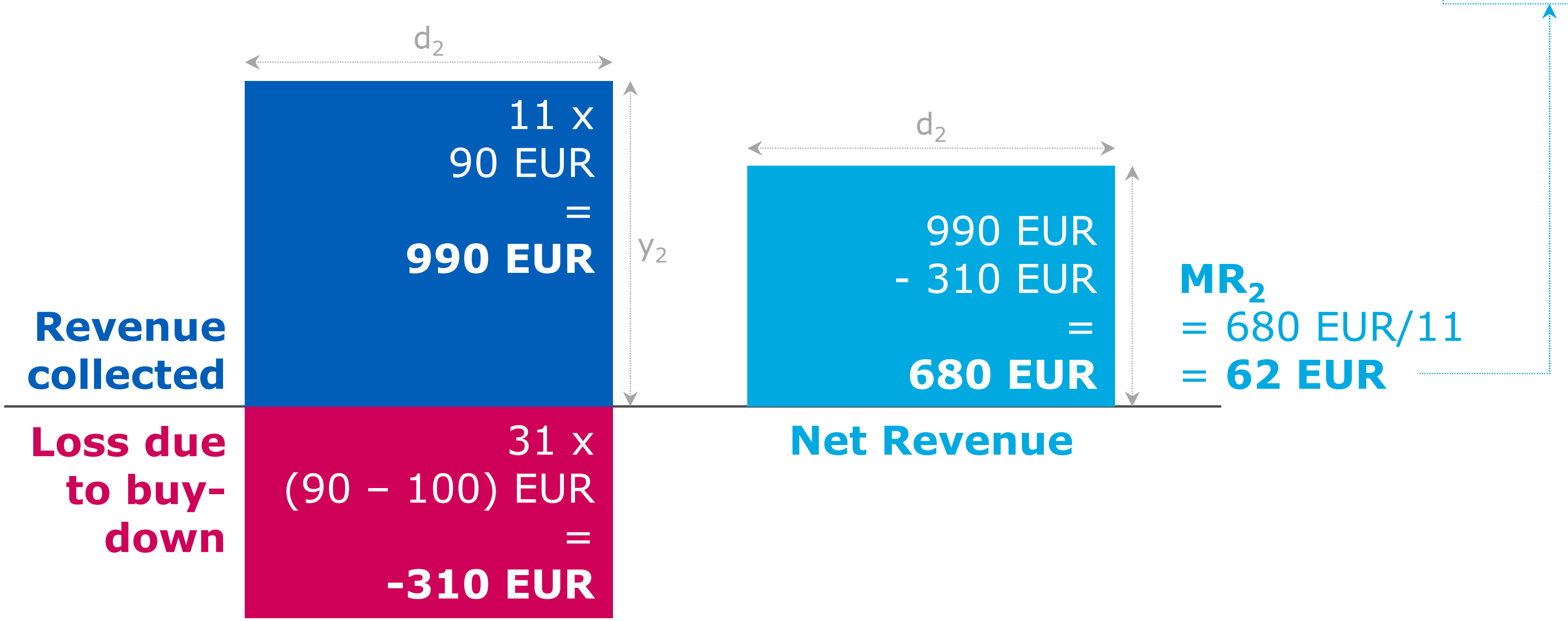
Confidential RESTRICTED



Confidential
RESTRICTED

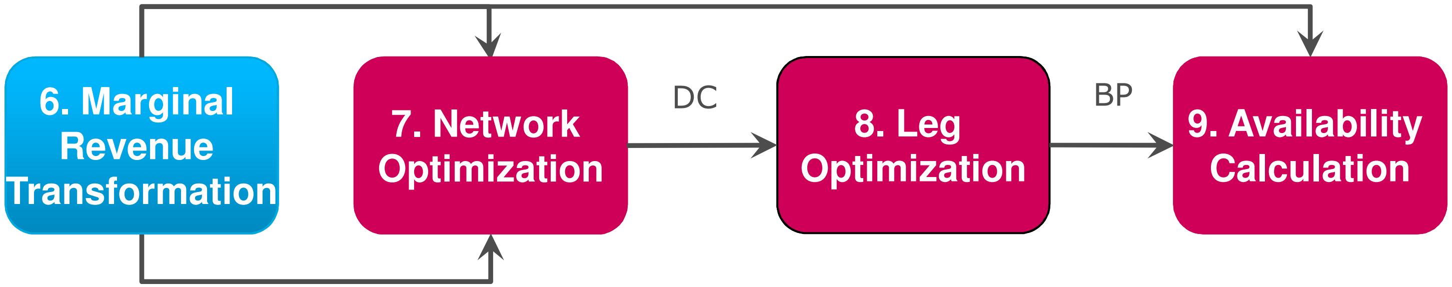
Marginal Revenue Transformation

Class	Yield y_i	Class dem. d_i	Total dem. Q_i	Total rev. TR_i	Marginal rev. MR_i
W	**100 EUR**	31	31	3,100 EUR	**100 EUR**
K	**90 EUR**	**11**	**42**	3,780 EUR	**62 EUR**



Marginal Revenue Transformation

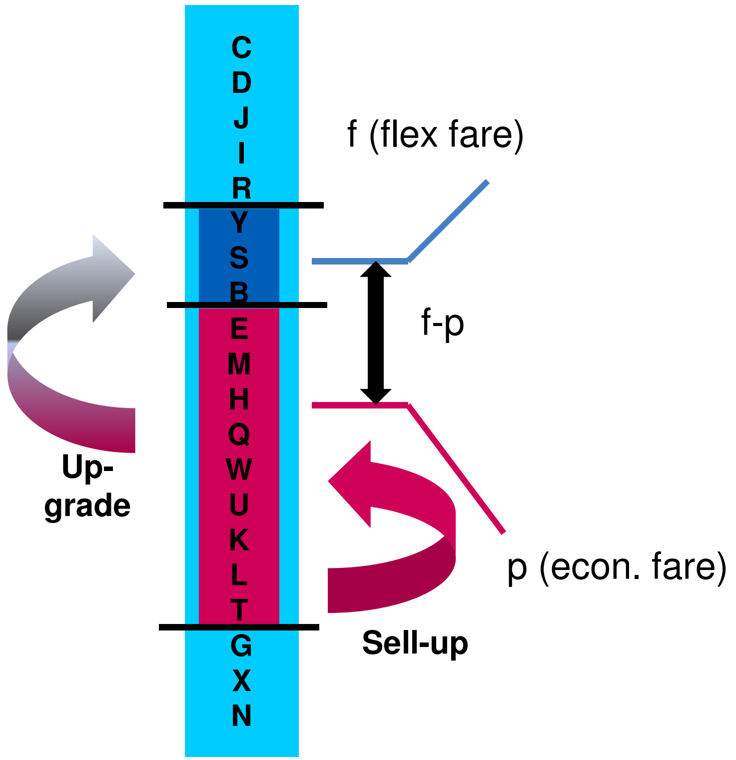
Adjusted yields (traffic flow)



Independent class demand
(traffic flow)

- Marginal Revenue Transformation calculates the marginal revenue and demand of accepting a booking considering the risk of buy-down.
- Marginal Revenue are calculated daily for all traffic flows, depdates and days to departure.
- The adjusted fares are used in both optimization and availability calculation.
- This ensures the right controls apply consistently in the real-time availability decision.

Fare Families from a RM perspective

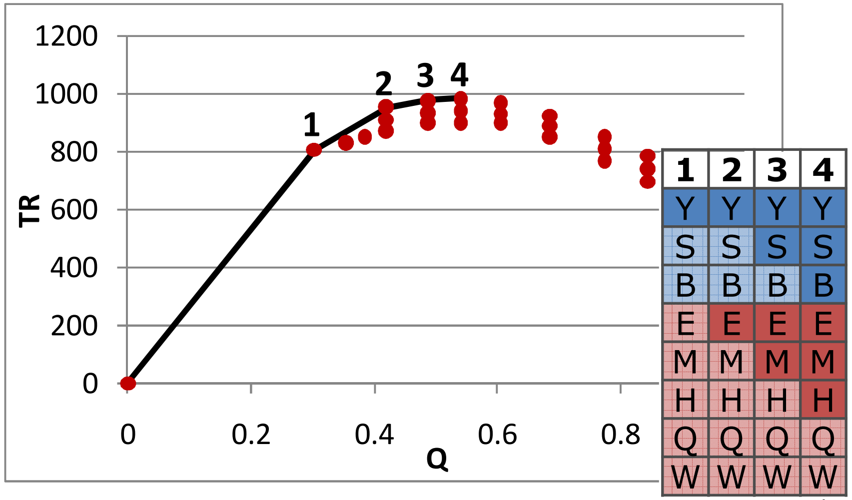


- Definition:
  - The fare structure consists of two or more fare families. Each fare family has the same set of restrictions (within a family price is the only difference).
  - Shown are the open fares in flex (f) and the open fare in economy (p). f-p is the up-grade level.
- RM objective: max total revenue requires:
  - Forecasting: Demand model that for any policy (p,f) predicts the demand in economy and flex.
  - Optimization: Determine the set of efficient policies (p,f) by path. The ordering sequence of the policies. Bid-price for the legs.

Confidential
RESTRICTED

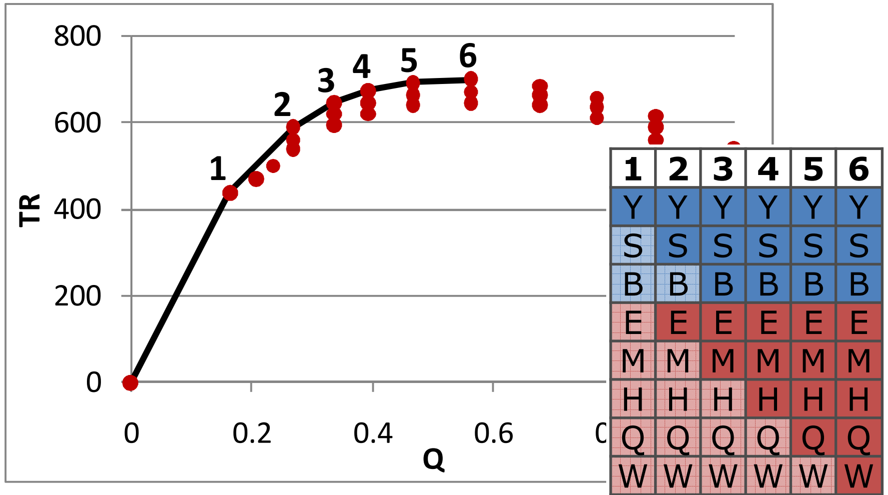
Application Fare Families
- Convex hulls by time frame

DBD=1

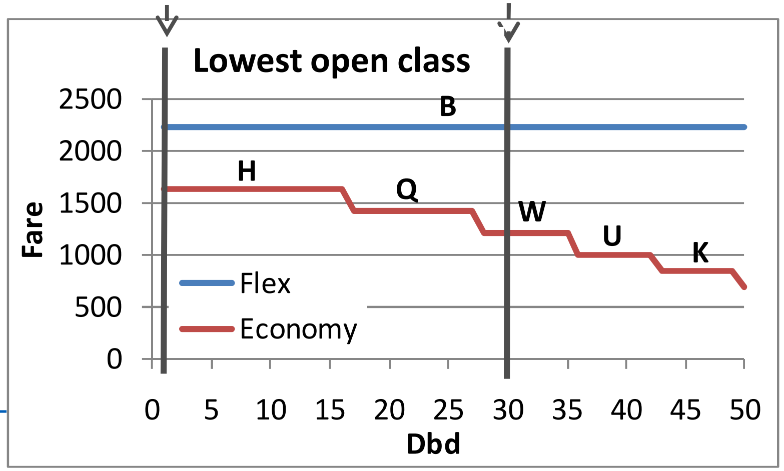


1	2	3	4
Y	Y	Y	Y
S	S	S	S
B	B	B	B
E	E	E	E
M	M	M	M
H	H	H	H
Q	Q	Q	Q
W	W	W	W

DBD=30



1	2	3	4	5	6
Y	Y	Y	Y	Y	Y
S	S	S	S	S	S
B	B	B	B	B	B
E	E	E	E	E	E
M	M	M	M	M	M
H	H	H	H	H	H
Q	Q	Q	Q	Q	Q
W	W	W	W	W	W



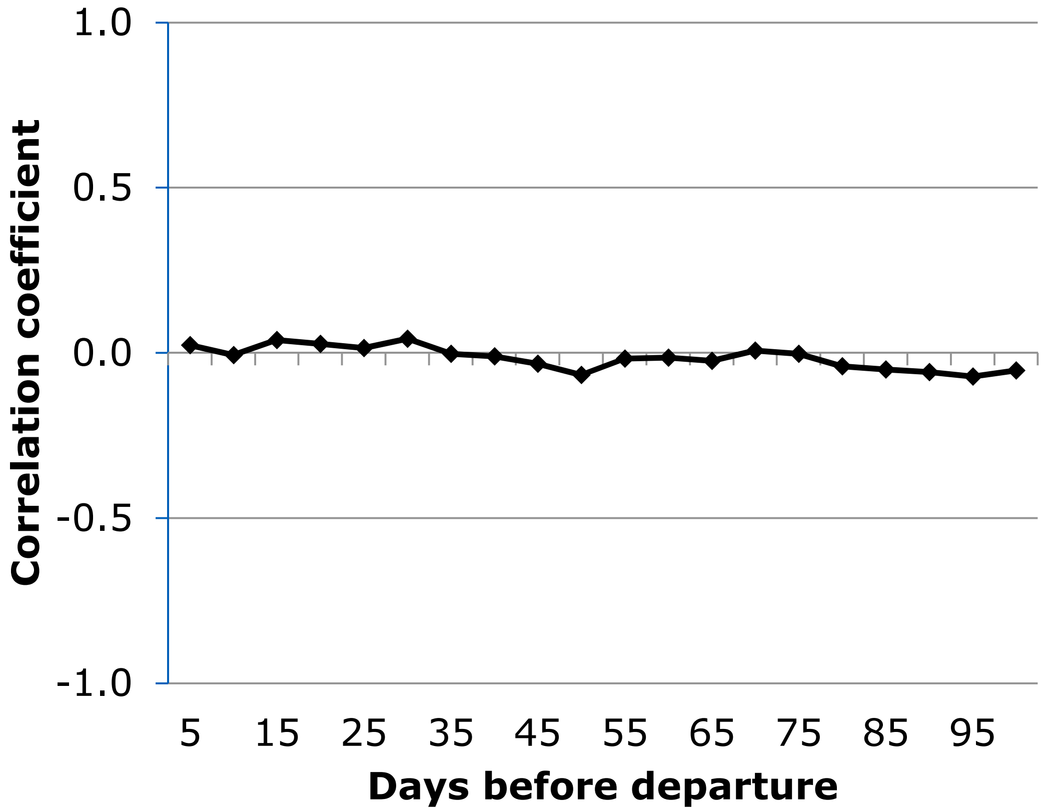
QATAR AIRWAYS القطرية | amadeus

7 Research on adaptive forecast

Should inventory be included in the model?

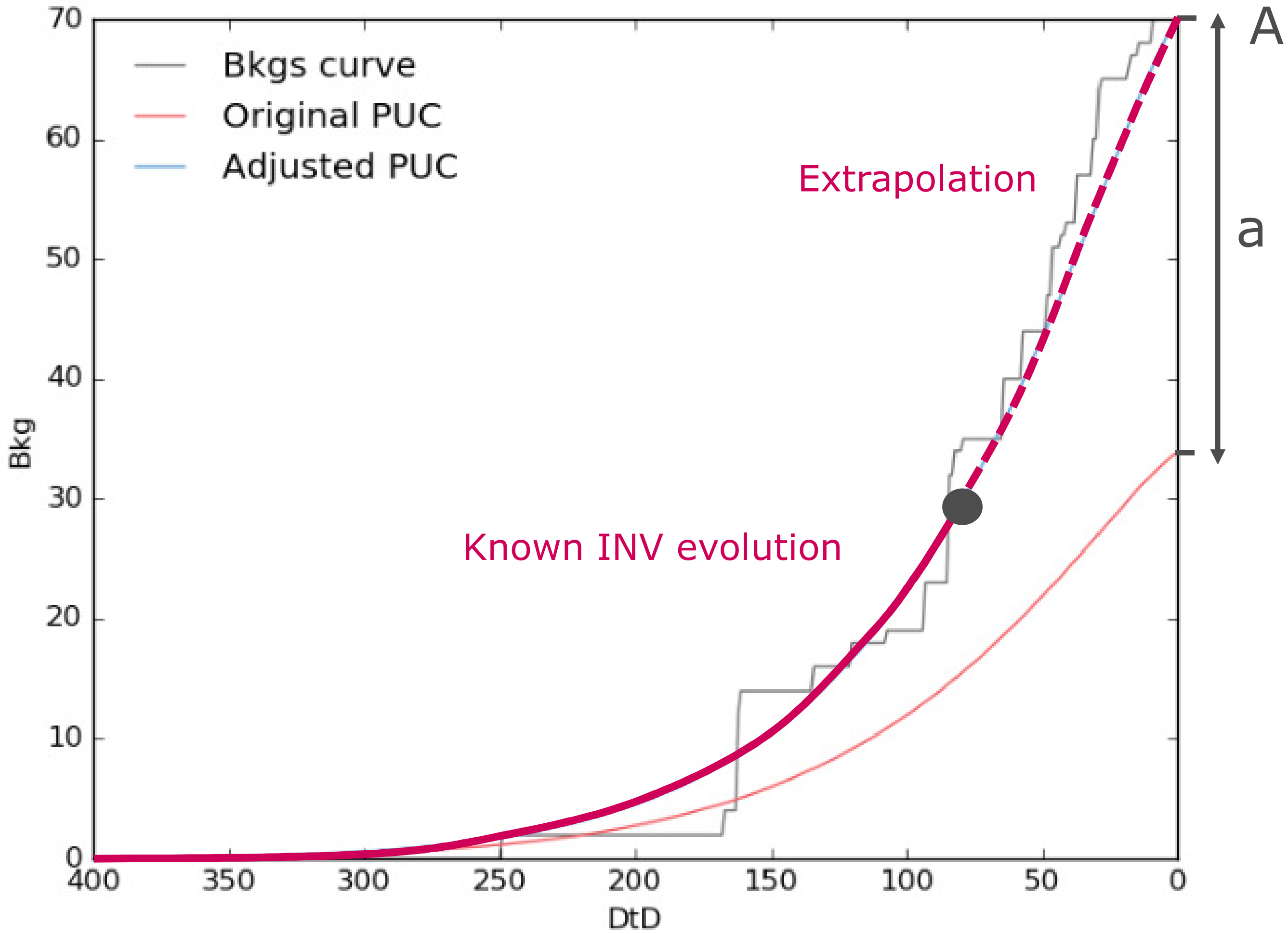
Correlation coefficient between current and future bookings

- **Correlation plot below bet8ween Corr[INV(dtd), INV(0)-INV(dtd)] ~ 0** suggests that INV does not provide any information of remaining demand.



Not the full story!

We know more: INV evolution

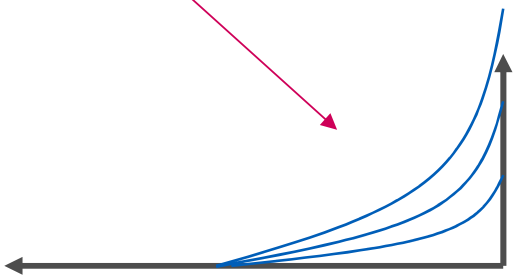


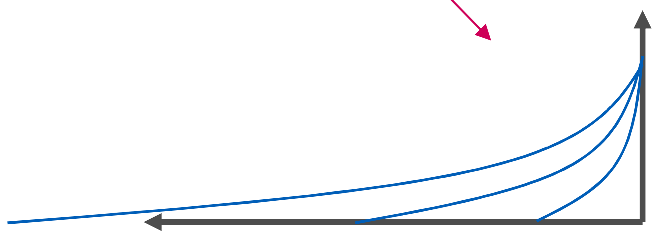
Adaptive Forecasts

Adjusted Forecast - Correction

$$fcst_{r,c}(\boldsymbol{p}, dep, dbd|pol = \{c\}) = vol(\boldsymbol{p}, dep, dbd) \times MNL(\boldsymbol{f}, \boldsymbol{\beta}) =$$

$$a \times gm_r \ w_k s_m \left(1 + a_1 \sin\left(2\pi \frac{doy}{365}\right) + a_2 \cos\left(2\pi \frac{doy}{365}\right)\right) e^{-\gamma \cdot \frac{dtd}{dtd_c}} (1 - e^{-\gamma}) MNL(\boldsymbol{f}, \boldsymbol{\beta})$$

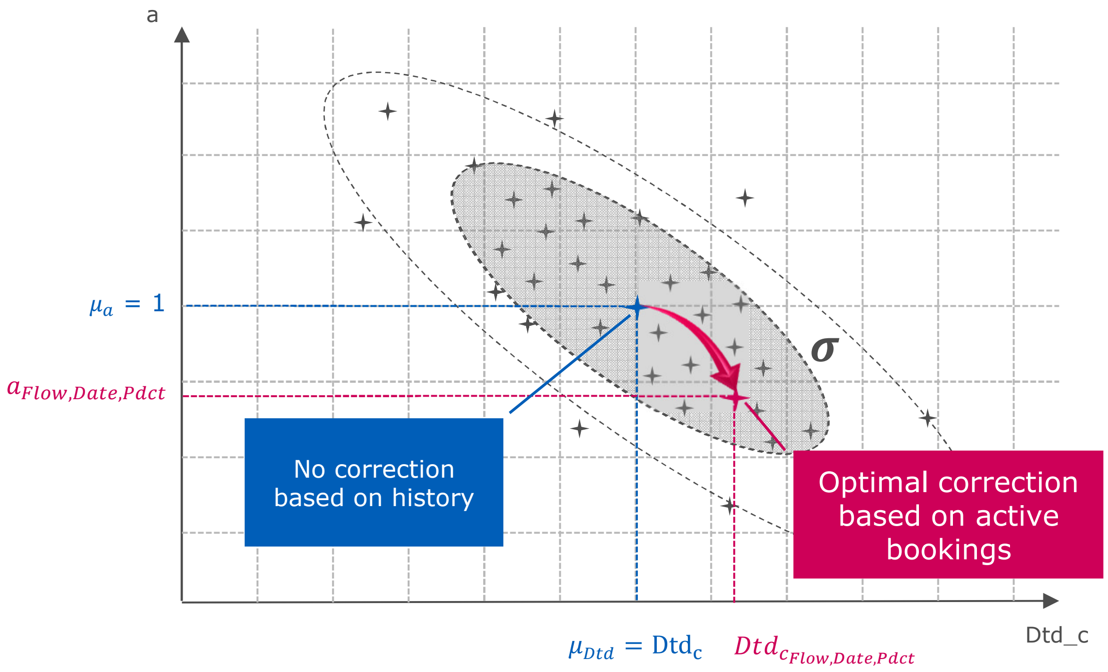




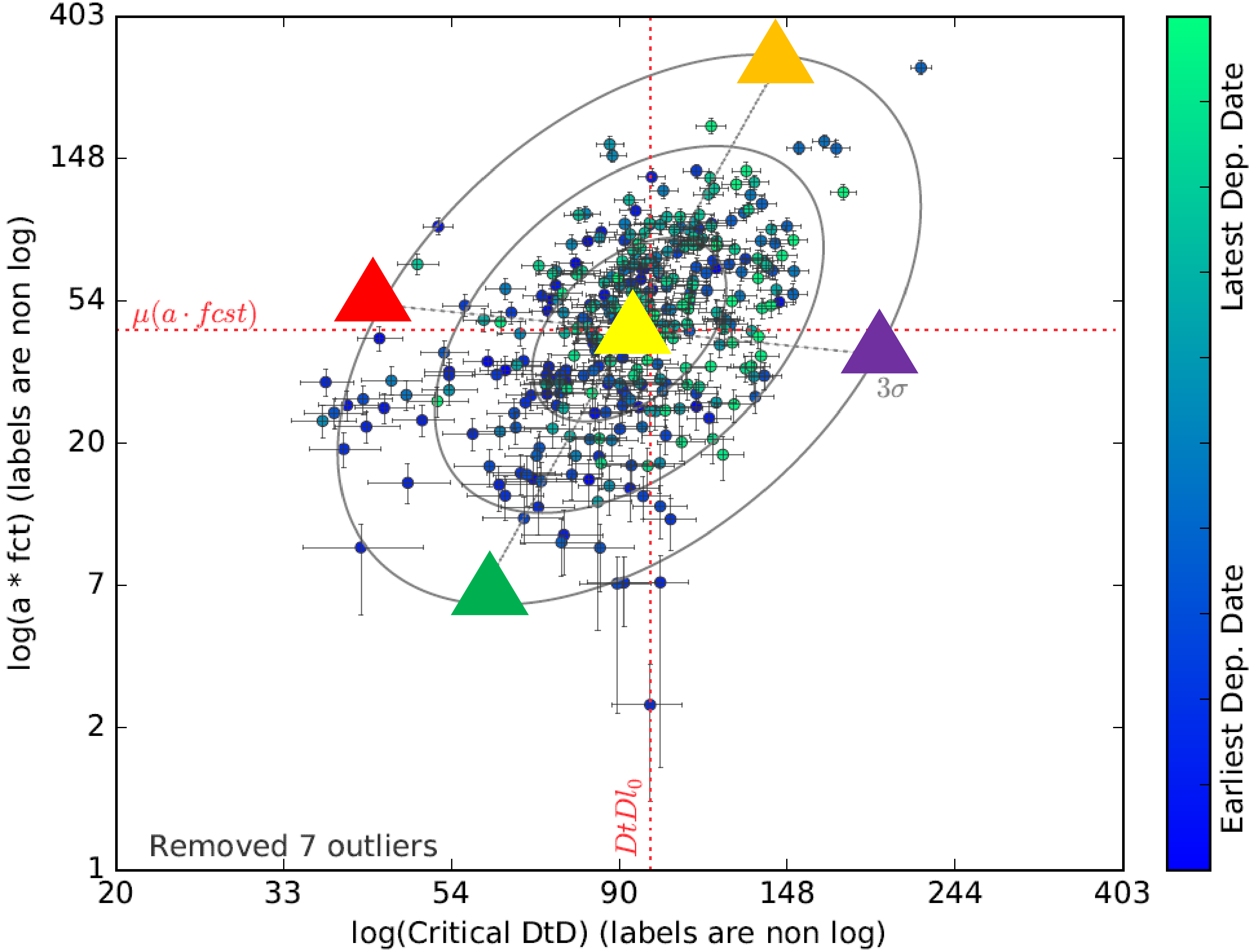
- Adaptive forecast constructed from existing forecast using linear affinities
- Two scale parameters:
  - Volume scale (a)
  - Booking horizon scale: (dtd critical)

Adaptive Forecasts

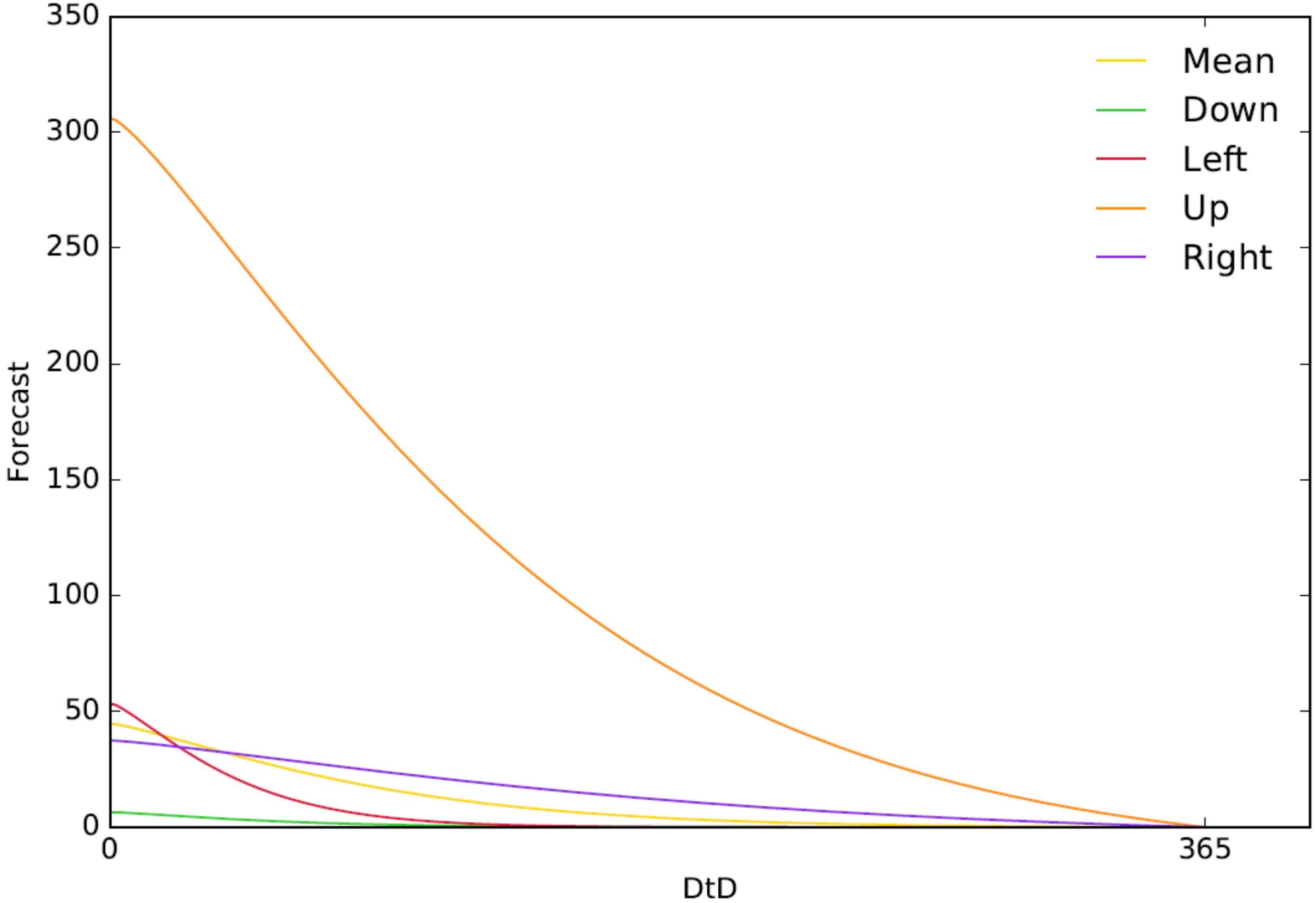
Effect on an active Flow Date



Correlation graph

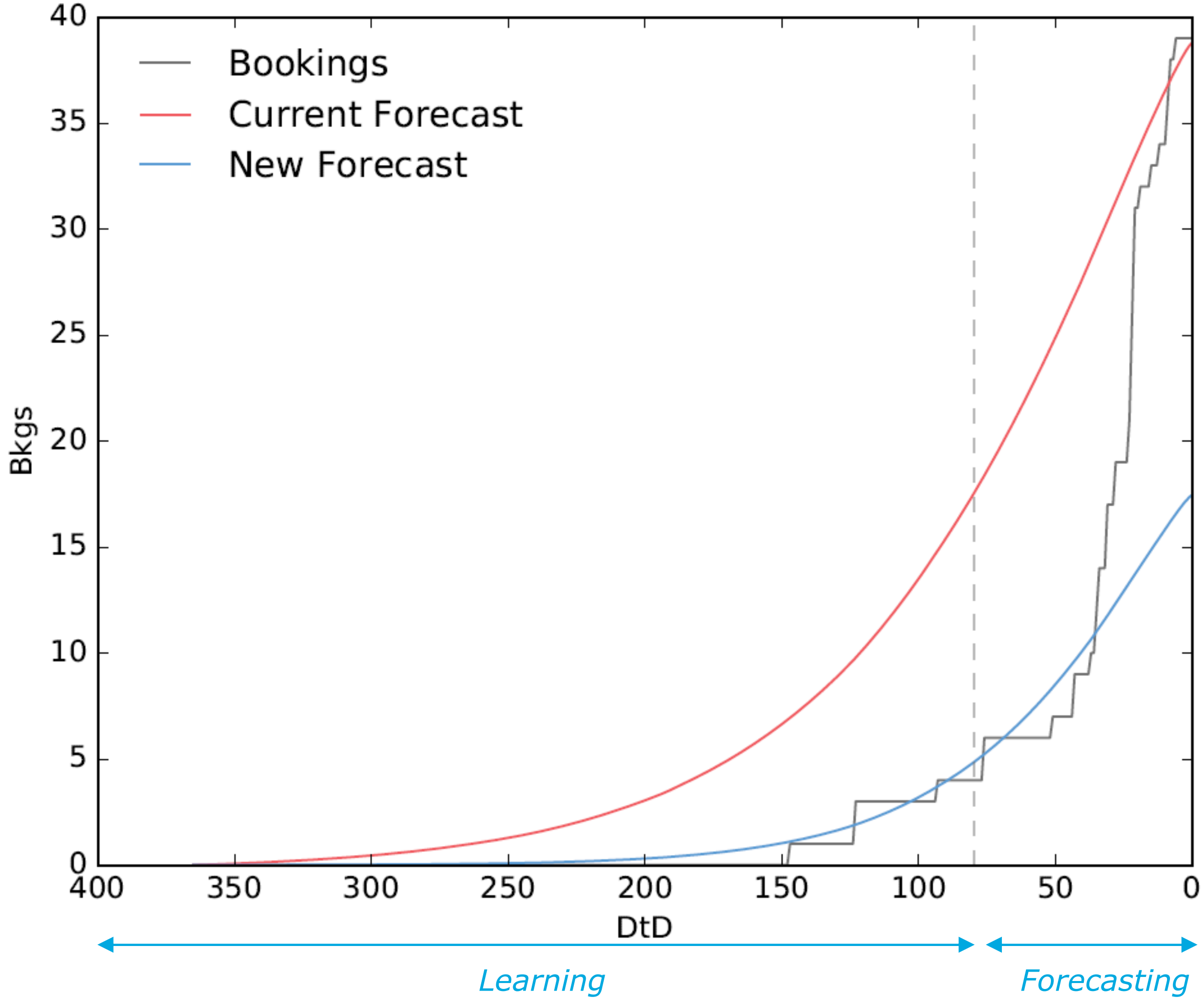


Historical PU Curves - Examples



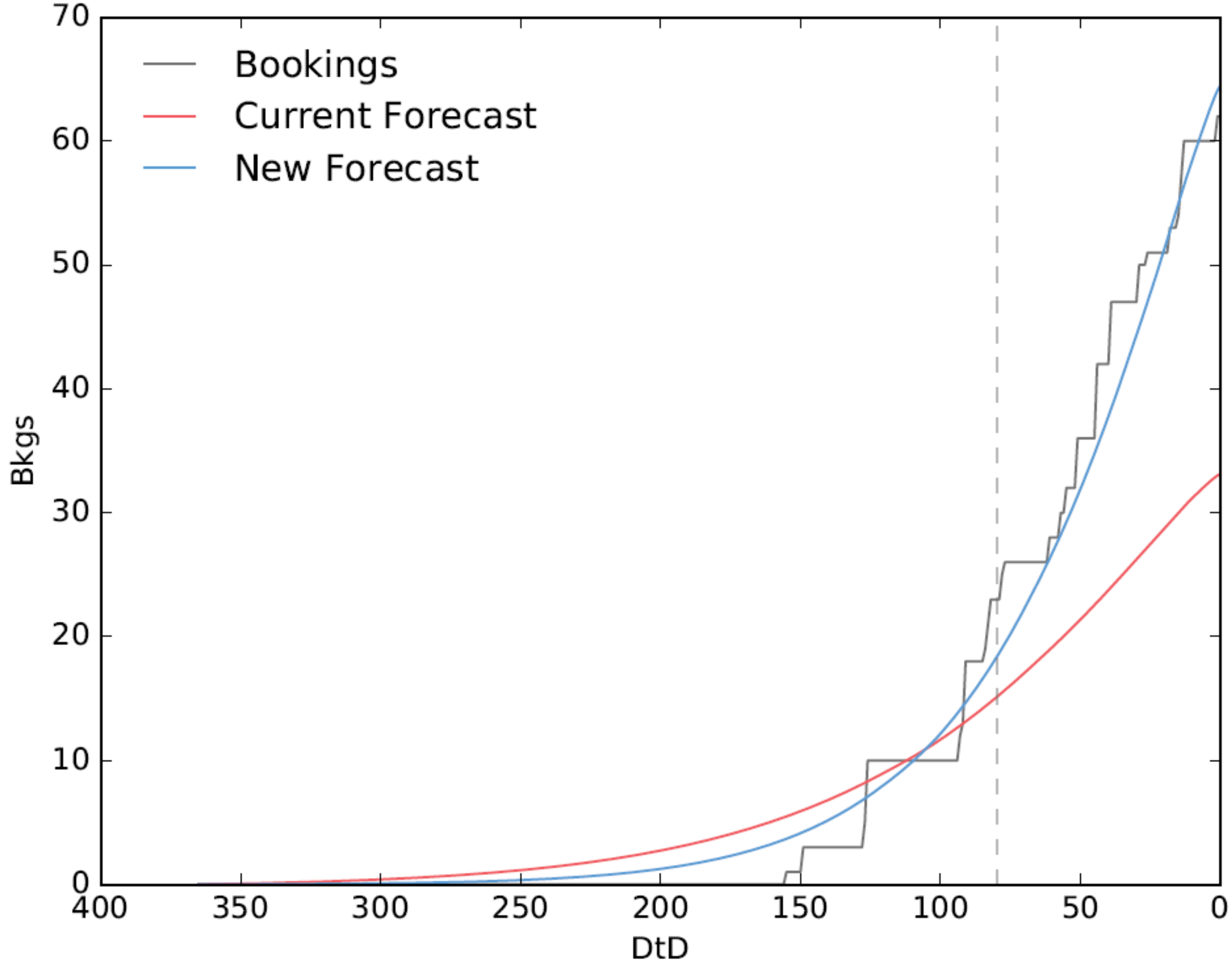
Adaptive Forecasting - Example cases

Not enough learning, bad results



Adaptive Forecasting - Example cases

Good learning

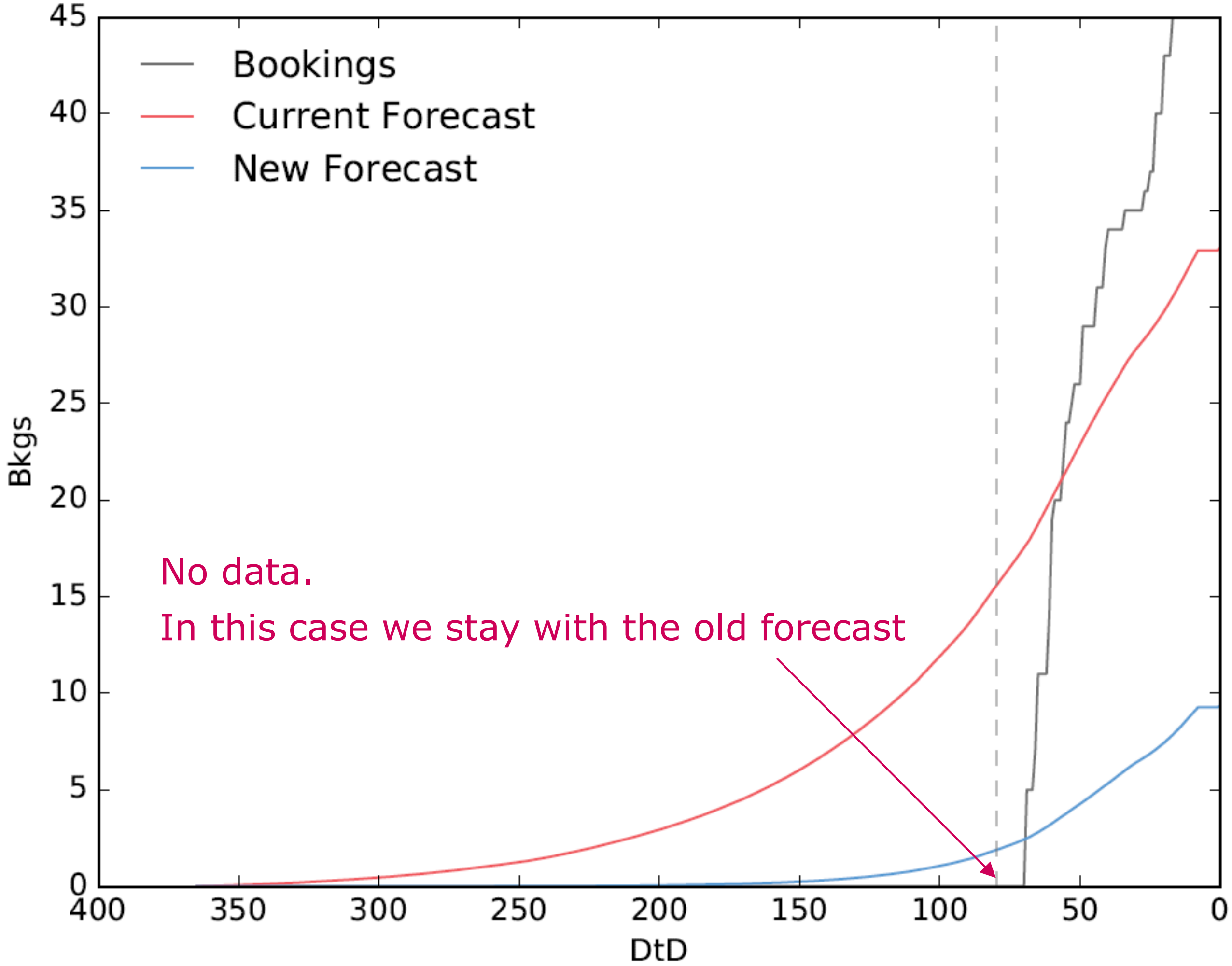


QATAR AIRWAYS القطرية | amadeus

Adaptive Forecasting - Example cases

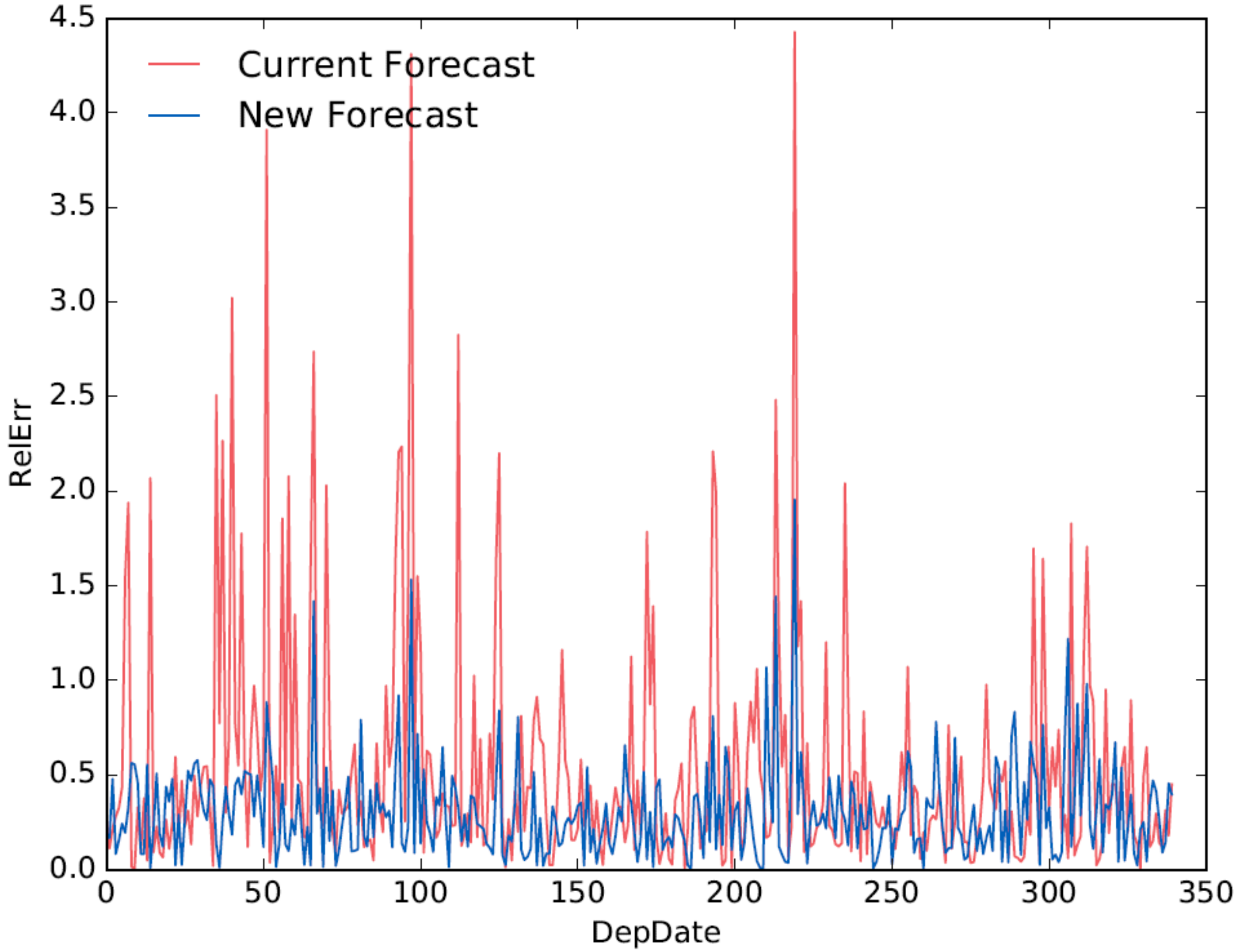
Confidential RESTRICTED

We can't always predict - No Data



Adaptive Forecasting

Promising results.



- **Forecast quality measured as:**

$$\mathrm{Err} = \frac{Fcst - Bkd}{Bkd}$$

- **Current forecast: Err 0.54**
- **Adaptive forecast**: **Err 0.30**

- Data has been cleaned so that active data is without cancellations. Hence the benefit will be reduced in the real world.

Confidential RESTRICTED

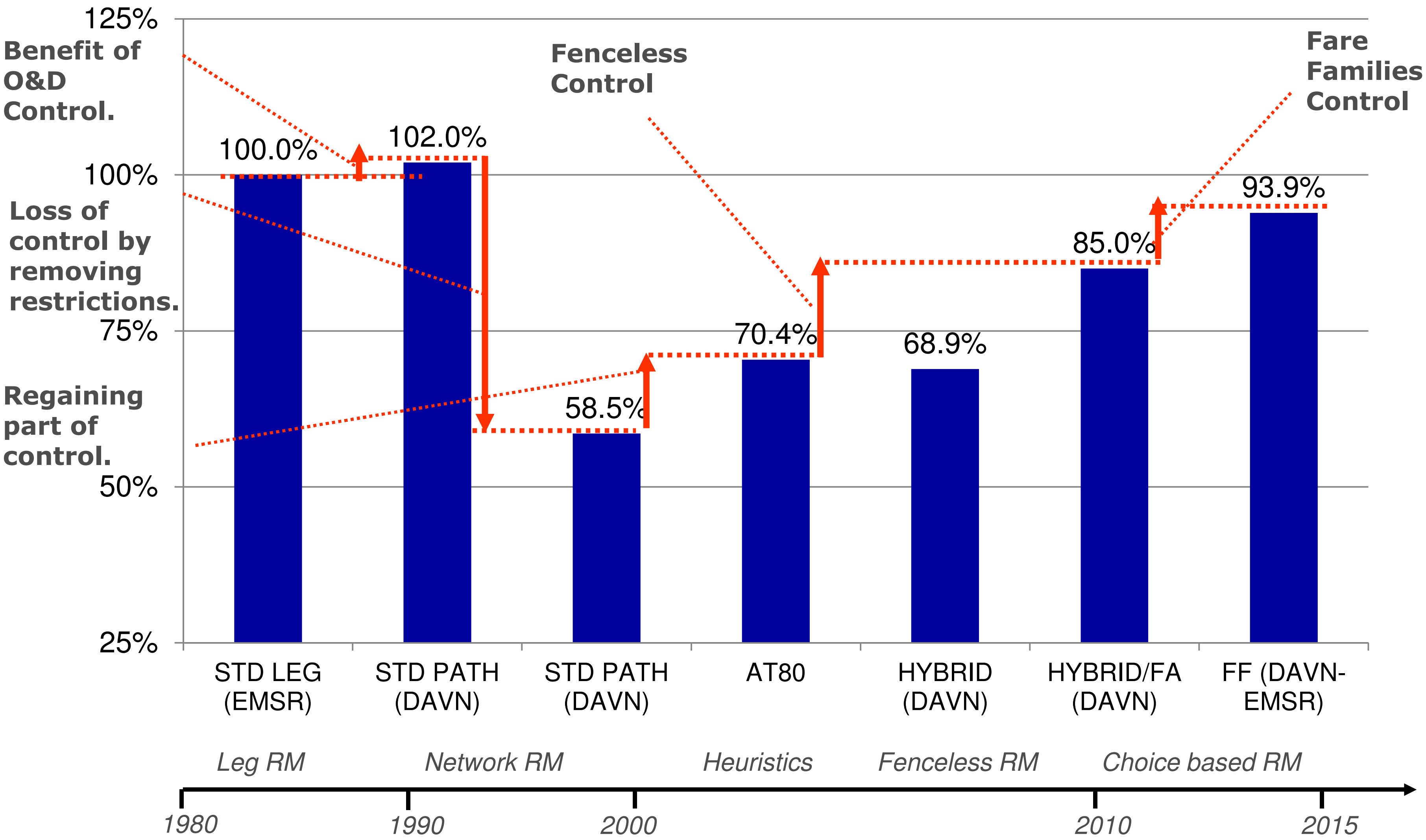
QATAR AIRWAYS | amadeus

8 Conclusion

Confidential RESTRICTED

PODS Benchmark studies*)

Relative revenue performance



*) Cf. Hopperstad Consulting Report Dec 2012.

Comparison between forecasting models 1

Confidential RESTRICTED

	Altéa O&D forecast model	Traditional forecast models	PODS
Volume forecast by DOW, Seasonality, Special day, Trends	**Yes** (estimation done in consistently)	**Yes** (estimation done in multiple steps)	**No** (Only seasonality possible)
Un-constraining	**Yes** (Done implicitly by minimize constrained forecast error)	**Yes** (Done explicitly by un-constraining obs. Cannot be observed/verified	**Yes** (Done explicitly by un-constraining obs. Cannot be observed/verified
Forecasts type	**Net-forecasts** (True survived bookings from PNRs)	**Gross-forecasts** (Most advanced. Canc method 3 from PODS)	**Net and Gross-forecasts** (Currently testing)
Use live data in estimation	**No** (Only departed observations used)	**Yes/No** (Can use live data but issue with canc.)	**Yes/No** (Available, but currently not used)
Cancellations	**Yes** (Forecast for no-show and cancellations)	**Yes** (Forecast for no-show and cancellations)	**Yes** (Currently testing)
Estimation methodology	**Parametric** Analytical using all historical data	**Parametric/Non-parametric** Update forecast parameters based on historical obs.	**Non-parametric** Update forecast parameters based on historical obs.
Calibration	**Automated** Self-calibrating every week using historic data	**Manual** Done at installation	**Automated** During warm-up

QATAR AIRWAYS القطرية | amadeus

Comparison between forecasting models 2

Confidential
RESTRICTED

	Altéa O&D forecast model	Traditional forecast models	PODS
O&D Traffic flow, O&D data collection	**Yes** (Booked, Yield, Taxes, Ava at the detailed level)	**(Yes)** (Limited access to Ava. and Tax at O&D level)	**(Yes)** Access to AVA. Tax excluded. All flows significant.
Supported fare structures	Fenced, Fenceless, **Fare Families**	Fenced, Fenceless, **Fare Families NOT supported**	Fenced, Fenceless, **Fare Families**
Monitoring forecast quality	**Yes** (Constructed to minimize constrained forecast error)	**No** (Fenced fare str. Or base demand for fenceless.)	**No** (Fenced fare str. Or base demand for fenceless.)
Estimation of the Volume at O&D traffic flow level	**Yes**	**Yes**	**Yes**
Estimation of the sell-up rate (Willingness to pay)	**Yes**	**No/(Yes)** (normally at segment level)	**No/(Yes)** (currently frat5 input, but can be estimated system-wide by product)
Robustness on sparse data	**Yes** (Due to all data being used in estimation; Parsimonious use of forecast parameters)	**No** (Unstable due to updates based on latest obs. to parms. at lowest level)	**No/Yes** Estimators for WTP by market tested were insufficient.
Use live inventory in forecast	**No** Correlation assumed to be zero	**No/Yes** (Live data can be included)	**No/Yes** (Available, but currently not used)

Key scientific elements of forecasting and price-sensitivity management

Confidential RESTRICTED

- The Altéa platform provides **high quality of data** to support the forecasting and WTP estimation.
- Holistic forecaster optimizes forecast accuracy by estimating all forecast parameters (un-constraining, volume, WTP, product preference) in a single process.
- The RM system is **automatically recalibrated weekly**.
- Manual inventory overrides are no longer required to limit the spiral down: **risk of buy-down is automatically measured** and converted into dynamic fare adjustments.
- Fare modifiers are automatically computed and applied in both network optimization and in the availability processor.
- User interventions at the traffic flow level allow adjustment of demand and WTP. The interventions trigger re-forecasting and re-optimization (including partial network) which are then propagated in real-time to the availability processor.

Confidential RESTRICTED

Thank you

You can follow us on:
AmadeusITGroup

amadeus.com/blog
amadeus.com

amadeus