Web Appendix for Interpretable Modeling of Retail Demand and Price Elasticity for Passenger Flights using Booking Data

Jan Felix Meyer [1], Göran Kauermann [1], and Michael Stanley Smith [2]

[1] Department of Statistics, Ludwigs-Maximilians-Universität München, Ludwigstr. 33, Munich, Germany

[2] Melbourne Business School, University of Melbourne, 200 Leceister Street, Melbourne, Australia

---

Address for correspondence: Göran Kauermann, Department of Statistics, Ludwigs-Maximilians-Universität München Ludwigstr. 33, 80539 Munich, Germany.

E-mail: `goeran.kauermann@stat.uni-muenchen.de`.

Phone: (+49) 89 2180 2220.

Contents

1 Revenue Management Literature Review

Airlines' revenue management systems handle thousands of transactions per second, and sell decisions are due in milliseconds, so that no current revenue management system works in real time. Therefore, accurate estimates of demand and price elasticity are essential to precompute control values to maximise revenue from of its seat inventory (McGill and Van Ryzin, 1999). Early examples of price elasticity estimation include Jung and Fujii (1979), Oum et al. (1992), Brons et al. (2002), and Kremers et al. (2002), while Granados et al. (2012) study differences in price elasticities for flights due to distribution channel using a log-linear model. In the revenue management literature, multinomial discrete choice models of customer selection between available booking classes, cabin classes and/or flights times are popular; for examples, see Vulcano et al. (2010), Vulcano et al. (2012), and Dai et al. (2014). In contrast, we do not model individual customer choice, but model daily booking counts for a given flight and cabin class. Any remaining dependence in these counts due to customer choice between different flights departing on the same day is instead captured by the copula model. A number of authors also combine nonhomogeneous Poisson processes for the arrival of potential customers, with product choice models (Balaiyan et al., 2019). A 'no-buy' option is included in the choice set to accommodate potential customers who do not buy a ticket; see Vulcano et al. (2012), Besbes and Zeevi (2015), and Van Ryzin and Vulcano (2014). However, our model differs from this because we model the realized booking process — i.e. the bookings that are actually made — rather than the arrival of potential customers, including those who do not make a booking.

Lo et al. (2015) and Li et al. (2014) also stress the importance of accounting for price endogeneity in models of demand in the airline industry, as do Mumbower et al. (2014) and Lurkin et al. (2017) who also make use of instrumental variables. Similarly, a number of other authors have also considered latent segmentation of customers when estimating demand and/or price elasticity using choice and other models; for example, see Teichert et al. (2008), Wen and Lai (2010), Martinez-Garcia and Royo-Vela (2010), Vij and Walker (2014), and Feldman and Topaloglu (2015). However, our study is the first of which we are aware that identifies such a rich latent segmentation of airline passenger bookings using a mixture-of-experts style model calibrated with a large disaggregate dataset. A similar modelling approach to our mixture-of-expert model is Li et al. (2014). The authors analyze strategic behaviour by a mixture of myopic and strategic customers, a special case of our model with two latent segments and no differentiation between the arrival time of strategic customers. Reviewing the statistical literature for two fundamental techniques accounting for unobserved heterogeneity, Sfeir et al. (2021) compare latent class with mixed logit models. Besides the advantages of latent class models, having fewer assumptions about the mixture distribution, being interpretable (as their mixture component typically depends on covariates — in our case, time to departure, departure time, and booking day of the week), and the correlation between the mixture component and the segment-specific variables and estimated elasticities are implicit in the model (mixed logit models need to assume a joint distribution for both components), the author mentions that latent class models may oversimplify the unobserved heterogeneity if the number of classes is small. To ensure that our model does not oversimplify the unobserved heterogeneity, up to 7 passenger segments are analyzed for each departure day.

Last, Wen and Chen (2017) account for the impact of the days to departure at booking on demand as a smooth nonlinear function, whereas Lurkin et al. (2017) do so for the flight departure time. Both papers employ parametric function bases constructed from low order Fourier terms. In contrast, following the statistical literature (Wood, 2017), we model both nonlinear effects using splines. These are more flexible and allow for data-driven levels of smoothing. Moreover, these two nonlinear effects are estimated for both the booking intensity and the mixture probabilities in the MNL.

Within Table A.1, we summarizes the main features (sample size of data, usage of covariates, model-type, handling of endogeneity (Endo.), and usage of segmentation (Seg.)) of prior studies of passenger flight retail demand and price elasticity that are closest to ours.

Table A.1: Comparison of relevant prior literature on modelling retail demand for passenger flights

Author	Data	Covariates	Model	Endo.	Seg.
Dai et al. (2014)	Booking, ticketing, and availability data from 3 airlines between 2011 and 2012 (n = 748,076).	[1] Ticket price, [2] Departure Time, [3] Ticket change fee, [4] Milage gain, [5] Carrier, [6] Booking Time, [7] Booking Channel	MNL, Nested Logit, Mixed Logit	no	no

Mumbower et al. (2014)	JetBlue Webbot data for transcontinental flights between 2 and 22 September 2010 over a 28-day booking horizon (n = 7,522).	[1] Ticket price, [2] Departure day of week, [3] Departure Time, [4] Days to departure at booking, [5] Booking day of week, [6] Virgin America promotions, [7] Labor Day indicator	Ordinary Least Squares, Two Staged Least Squares	yes	no
Fiig et al. (2014)	Bookings at 22 selected traffic flows from Scandinavian Airlines (n = 7,780).	[1] Ticket price, [2] Departure day of week, [3] Departure Time, [4] Days to departure at booking, [5] Recurring special periods, [6] Departure Date	Nonlinear regression (multiplicative)	no	no
Vulcano et al. (2012)	Booking data from last 7 selling days for 11 Monday flights from January to March of 2004.	[1] 11 Products (fare-classes) with different fare-values, [2] 7 Booking Periods (each 24 hours), [3] 2 daily flights, [4] Market share	MNL	no	no

Teichert et al. (2008)	Stated preference survey data from frequent flyer passengers traveling on 11 European short-haul routes (n = 5,829).	[1] Product characteristics: compartment (business, economy), [2] Stated preferences: scheduled frequency, price, fare flexibility, punctuality, catering, ground service, [3] Behavioral and socio-demographic variables: gender, age, education level, profession, flying frequency	Latent Class	no	yes
Lurkin et al. (2017)	Ticket data purchased through travel agencies worldwide as collected by the Airlines Reporting Corporation (ARC) ($n = 10,034,935$).	[1] Itinerary Information (Departure Time, Travel Time, Equipment, Number of connections, Direct Flight Indicator), [2] Price (Average high yield fare, Average low yield fare), [3] Marketing relationships (Codeshare, Interline, Online), [4] Carrier preference	MNL	yes	no
Present Study	Retail bookings and flight data up to 120 days to departure (n = 1,333,712).	[1] Ticket price, [2] Departure day of week, [3] Departure Time, [4] Days to departure at booking, [5] Booking day of week	our model	yes	yes

2 Price Model

The initial step in our estimation is the construction of the residual $\hat{\xi}$ to accommodate the missing exogeneity of price. Table A.3 provides estimates of the linear

Component	Estimate ($\hat{\theta}_j$)	Std. Error	% Change
Intercept	4.6204	0.0014	–
log(IV)	0.1052	0.0003	11.09%
BDAY = Mon	-0.0148	0.0017	-1.47%
BDAY = Tue	-0.0170	0.0017	-1.69%
BDAY = Wed	-0.0141	0.0017	-1.40%
BDAY = Thr	-0.0127	0.0017	-1.26%
BDAY = Fri	-0.0054	0.0017	-0.54%
BDAY = Sat	0.0045	0.0019	0.45%

Table A.3: Linear parameter estimates for the model for PRICE at Eqn. (3.7) fitted to bookings departing on Thursday. The point estimate and the standard error are reported, along with the effect on PRICE of increasing each covariate by 1 unit, which is given by $\exp(\hat{\theta}_j) - 1$.

coefficients $\theta_0, \ldots \theta_7$ of model 3.7 and their impact on PRICE. The baseline for the BDAY dummy variable is Sunday, and the remaining weekday dummy variables have significant relationships with PRICE. There is a slight discount for tickets booked on weekdays of between 0.63% and 1.84%, compared to those booked on the weekends. The instrumental variable IV has a significant positive coefficient, with a z-value of $0.1052/0.0003 = 350.7$ for the null hypothesis that $\theta_1 = 0$; suggesting that the logarithm of bid-price is a strong instrument.

Figure A.1 plots the estimates of the smooth functions f_0 and f_1, along with 99%

confidence bands, constructed as in Marra and Wood (2012). The estimate of f_0 shows that ticket prices tend to increase closer to departure. Turning to the estimate of f_1, it can be seen that ticket prices tend to peak for flights departing at around 08:00 and 18:00. These are the morning and evening peak demand periods, and this increase is consistent with the demand profile for flights on a busy short-haul route.

3 Demand Model with Two Segments

Table A.4 gives the estimates of the linear coefficients, both excluding (Model I) and including (Model II) the residuals $\hat{\xi}$ from the price model. That is Model I ignores the missing exogeneity of price while Model II takes this into account through the above instrumental variable approach. We find a significant coefficients of $\hat{\xi}$, with z-

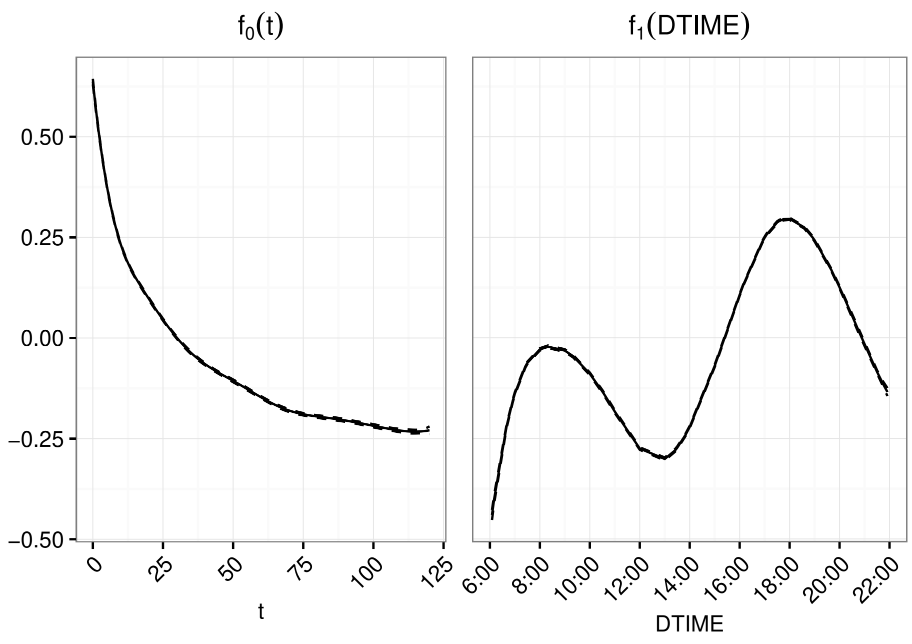

Figure A.1: Estimates of f_0 (left panel) and f_1 (right panel) from fitting the price model at Eqn. (3.7) to bookings on flights that depart on Thursday. The dashed lines are 99% confidence bands, which are tight.

Model	Segment / Component	BL $\lambda_{\mathrm{BL}}(t)$	$k=1$ $\pi_k(t)$	$k=2$
I			*Segment Adjustment Coefficients*	
	PRICE		**-0.0009** (0.0002)	**-0.0336** (0.0019)
		Baseline Coefficients	*Log-odds Coefficients*	
	Intercept	-1.1213 (0.0886)	-3.8781 (0.1450)	–
	BDAY = Mon	0.7753 (0.0743)	3.9991 (0.1437)	–
	BDAY = Tue	0.9768 (0.0760)	3.9786 (0.1540)	–
	BDAY = Wed	1.2515 (0.0806)	3.4608 (0.1395)	–
	BDAY = Thr	1.0012 (0.0877)	3.9168 (0.1482)	–
	BDAY = Fri	0.6714 (0.0848)	4.3488 (0.1605)	–
	BDAY = Sat	-0.3964 (0.0558)	0.3479 (0.1175)	–
II			*Segment Adjustment Coefficients*	
	PRICE		**0.0008** (0.0005)	**-0.0641** (0.0104)
	$\widehat{\xi}$		-0.0032 (0.0004)	0.0045 (0.0039)
		Baseline Coefficients	*Log-odds Coefficients*	
	Intercept	-0.9589 (0.5145)	-1.5856 (1.1263)	–
	BDAY = Mon	1.1680 (0.3021)	2.5242 (0.4329)	–
	BDAY = Tue	1.1535 (0.3879)	2.6881 (0.6621)	–
	BDAY = Wed	1.0423 (0.3924)	3.2702 (0.5823)	–
	BDAY = Thr	1.0721 (0.3397)	2.5534 (0.4931)	–
	BDAY = Fri	0.9795 (0.3058)	2.5480 (0.4754)	–
	BDAY = Sat	-0.4299 (0.0665)	-0.1150 (0.1905)	–

Table A.4: Parameter estimates and bootstrapped standard errors in parenthesize for $K = 2$ segments fitted to bookings on flights departing on Thursday. Results are given for models fit excluding (Model I) and including (Model II) the price model residuals $\hat{\xi}$.

statistics of $-0.0032/0.0004 = -8$ and $-0.0045/0.0010 = -4.5$ clearly highlights the importance of controlling for endogeneity here. Turning to the segments coefficients in Model II, the estimates are $\hat{\alpha}_{1,1} = 0.0008$ and $\hat{\alpha}_{1,2} = -0.0641$, suggesting that the first segment (which we label segment 1) consists of price inelastic customers, whereas the second consists of customers who are more price sensitive. Given the nature of the busy short-haul route, it is likely that segment 1 corresponds to a high proportion of customers travelling for business purposes, whereas the second segment includes a higher proportion of leisure travellers who are more budget conscious. Comparing the estimates of the coefficients of PRICE for Models I and II shows that controlling for endogeneity excentuates the price elasticity for segment 2.

Recall that the reference category in the log-odds at Eqn. (3.4) is segment $K = 2$. Therefore, the estimates of $\beta^{(\pi)}_{2,1}, \ldots, \beta^{(\pi)}_{7,1}$ indicate the relative preference of customers in segment 1 for booking on different day types. The positive (and significant) coefficient values for the weekdays indicate that customers from segment 1 are more likely to make a booking on weekdays, rather than on Saturday or Sunday (where the latter is the baseline case for the BDAY dummies). This is consistent with the interpretation of customers in segment 1 booking flights for business purposes.

Figure A.2 plots the estimates of the smooth components $s^{(\pi)}_{0,1}(t), s^{(\pi)}_{1,1}(\text{DTIME})$ for the log-odds equation as well as the estimates of the smooth components $s^{(\lambda)}_{0}(t), s^{(\lambda)}_{1}(\text{DTIME})$ of the baseline booking along with 99% confidence intervals for Model I. As the right-hand panels show, the probability of a booking increases with time getting closer to the departure date. Other than that there is only a little variation in time and DTIME which, as we will see, is also due to the fact that the model with $K = 2$ customer segments is too simplistic and does not appropriately describe customers' behavior. Comparing the results of Model I with the estimates of Model II,

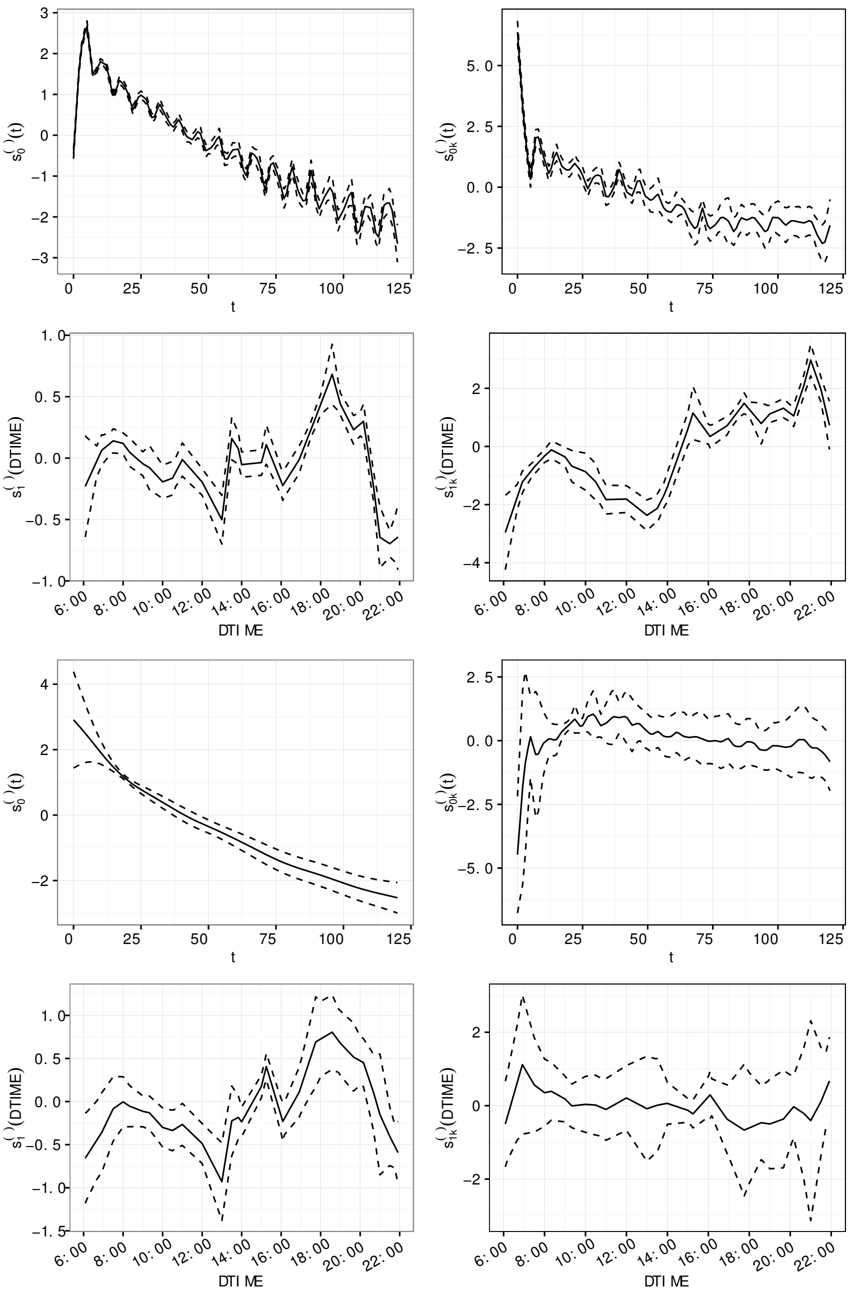

Figure A.2: For $K = 2$ segments, function estimates are given for models fit excluding (Model I, first two rows) and including (Model II, row 3 and 4) the price model residuals $\hat{\xi}$. The left-hand panels provide the function estimates for $s_0^{(\lambda)}(t)$ and $s_1^{(\lambda)}(\text{DTIME})$ in Eqn. (3.2) for bookings on flights that depart on Thursday. The right-hand side shows the estimates of $s_{0,1}^{(\pi)}(t)$ and $s_{1,1}^{(\pi)}(\text{DTIME})$ in Eqn. (3.4). The estimates are given by the solid line, while the dashed lines are 99% local confidence bands.

Figure A.2 shows that the smooth components are less erratic if controlling for price endogeneity. This is explained by the reduction in the models' degrees of freedom (134.42 for Model I and 94.73 for Model II). Though both models show similar results long before departure, i.e., overall booking intensity is relatively low and the mix of segments show increasing booking probability of the price-insensitive segment if going closer to departure, the most striking differences show during the week before departure. Here, $s_0^{(\lambda)}(t)$ no longer decreases whereas $s_{0,1}^{(\pi)}(t)$ indicates a decreasing booking probability of the price-insensitive segment. Additionally $s_{1,1}^{(\pi)}(\text{DTIME})$ is no longer significant proposing no segment specific booking probabilities with respect to departure time. As the interpretations of the smooth components from Model I and Model II make equal sense, i.e., Model I suggest that there is a general decline in booking intensity the week prior to departure where only the price-insensitive segment books whereas Model II depicts a steadily increase in booking intensity and a price-sensitive segment close to departure (last minute passengers only willing to travel if the price is cheap) this indecisiveness points towards the possibility of having at least an additional segment of price-sensitive passengers which Model I is not able to describe.

Figure A.3 plots $\bar{q}_1(t)$ and $\bar{q}_2(t)$, see (4.1) for Model II. The upper row shows that the proportion of customers in segment 1 — customers with demand patterns consistent with business travel — increases as the departure day gets closer. A strong weekly pattern due to the booking day type is also apparent. The bottom row of Figure A.3 $\bar{q}_1(\text{DTIME})$ and $\bar{q}_2(\text{DTIME})$ for Model II. We see that the proportion of customers in segment 1 increases during the peak periods during the morning and evening, which is also consistent with business travel.

Last, we estimate any over-dispersion in the Poisson model by computing the Pearson residuals

$$\varepsilon_{i,t} = \frac{Y_i(t) - E(Y_i(t))}{(\mathrm{Var}(Y_i(t))^{1/2}} = \frac{y_{i,t} - \lambda(\boldsymbol{x}_{i,t}, t; \boldsymbol{\theta})}{\lambda(\boldsymbol{x}_{i,t}, t; \boldsymbol{\theta})^{1/2}}.$$

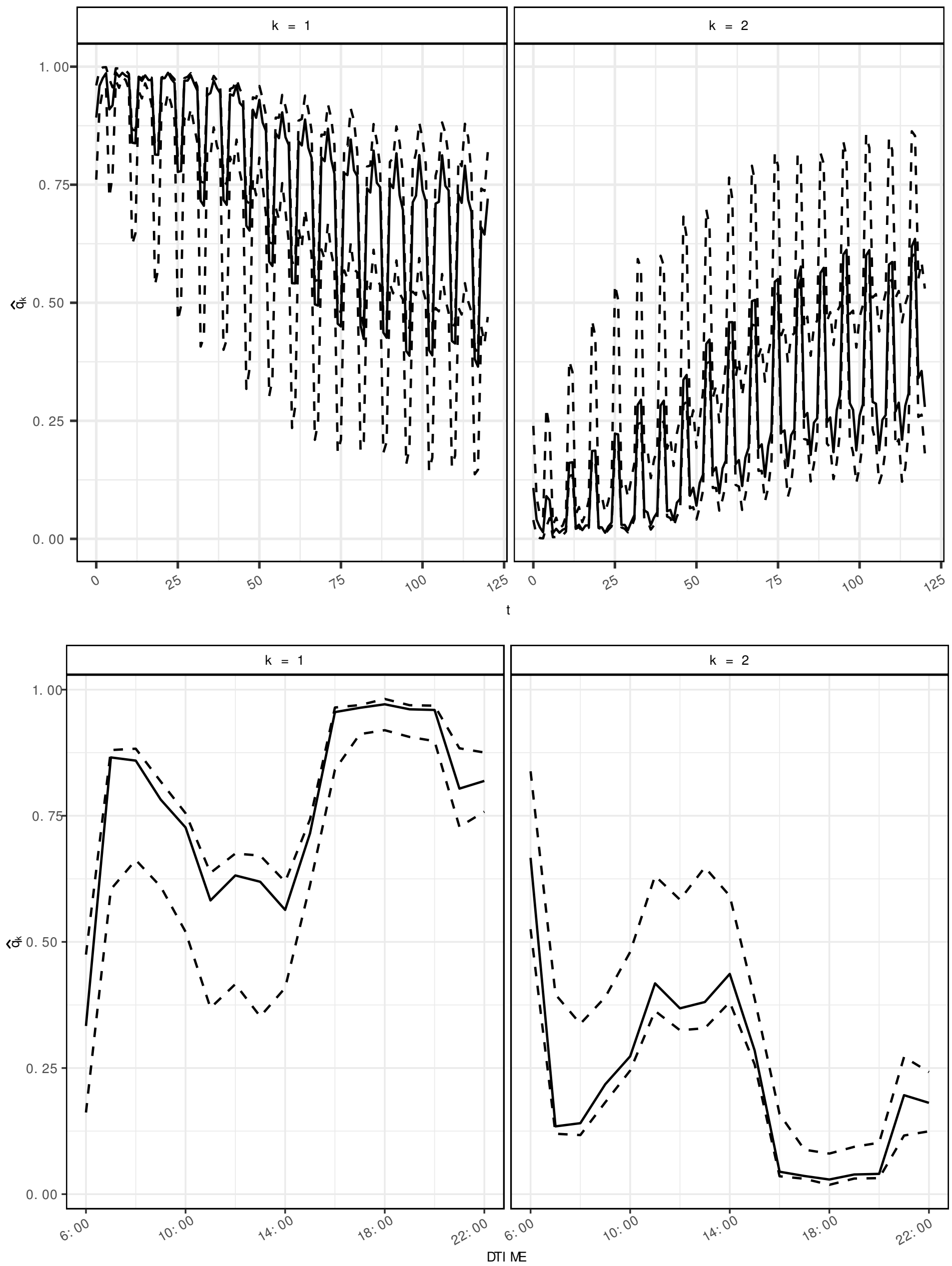

Figure A.3: Plot of the average segment proportion computed from the model fitted to booking on flights departing on Thursday and $K = 2$ (solid line) with 99% local bootstrapped confidence bands (dashed lines). Top row: within each panel, $\bar{q}_k(t)$ is plotted against days to departure t. Bottom row: within each panel, $\bar{q}_k(\mathrm{DTIME})$ is plotted against DTIME.

The mean of the squared residuals is 1.77, indicating only moderate over-dispersion to the Poisson model. We also investigated whether the squared residuals are related to the covariates, and also to the intensity, and we find no indication of structured heterogeneity.

4 Model Evaluation for intra-day dependence

So far we have treated bookings as independent, conditional on the covariates. However, dependence may exist between bookings made on the same day for flights departing on a given day, that is unaccounted for by the Poisson regression model. We call this 'intra-day dependence' in bookings, and to account for it we use a multivariate Gaussian copula model (Song, 2000) with the margins given by the Poisson regression models fitted above. Copulas models for discrete-valued responsesn have been used previously in the transportation sciences literature; for examples, see Bhat and Eluru (2009),Eluru et al. (2010), and Smith and Kauermann (2011). However, here our copula model needs to capture dependence between vectors that differ in length and composition for each observation, as we now discuss.

4.1 Gaussian Copula Model

In a copula model, dependence between the elements in a random vector of length m is captured by its 'copula function'. In practice, only vine or elliptical copulas currently are suitable for problems where $m \geq 3$ and pairwise dependence can vary between elements. Vine copulas can be difficult to specify (Dissmann et al., 2013), so that we instead use the Gaussian copula, which is the most popular elliptical copula. It has copula function

$$C_{\text{Ga}}(u_1, \ldots, u_m; \Gamma) = \Phi_m(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_m); \Gamma) = \Phi(y_1^\star, \ldots, y_m^\star; \Gamma)\,,$$

where $\Phi_m(\cdot;\Gamma)$ is the distribution function of a $N(\mathbf{0},\Gamma)$ density with Γ as correlation matrix, and Φ is the standard normal distribution function. The $m \times m$ correlation matrix Γ is the copula parameter that requires estimation. We define $y_l^\star = \Phi^{-1}(u_l)$ for $l = 1,\dots,m$ so that $(y_1^\star,\dots,y_m^\star) \sim N(0,\Gamma)$. As discussed in Danaher and Smith (2011), because the marginal distributions of the bookings are discrete-valued, we link the continuous Gaussian copula to the observed bookings through the constraint

$$\Phi^{-1}(\mathrm{Po}(y_l - 1;\lambda_l)) < y_l^\star < \Phi^{-1}(\mathrm{Po}(y_l;\lambda_l))\,,$$

with $\mathrm{Po}(\cdot;\lambda_l)$ as the distribution function from the fitted Poisson regression model from above.

Because on our route flights depart at 61 distinct times, we consider capturing intra-day dependence at the hourly resolution, with flights departing in hourly intervals from 06:00 to 22:00 (except for flights departing between [06:00,08:00) which we consider as a single interval because only a few flights depart prior to 07:00). This requires estimation of a 15-dimensional copula function $C_{\mathrm{Ga}}(u_1,\dots,u_{15};\Omega)$, where the parameter matrix $\Omega = \{\omega_{i,j}\}$ for $i = 1,\dots,15$ and $j = 1,\dots,15$. For example, element $\omega_{2,4}$ captures the dependence between flights departing in intervals [08:00,09:00) and [10:00,11:00). However, there are four complicating factors that make specifying the likelihood of such a copula model difficult for our data. For any given departure day d, (i) the number of flights scheduled to depart varies, (ii) the hours at which these flights depart varies, (iii) multiple flights can leave during the same hourly interval (particularly during peak periods), and (iv) a different set of flights can be open or closed for different booking days $t \in \{0,\dots,120\}$. Thus, the vector of booking counts for each departure day d and day to departure t, given as pair (d,t), can be considered to have 'ragged edges', because it differs both in size and composition of its elements. That is to say, in practice we do not have multiple observations on the 15-dimensional vector of bookings and hence direct application of a copula model to address intraday dependence is not possible.[1]

To account for these complications, we introduce the following notation. For each pair of values (d,t), let $K(d,t)$ denote the number of flights for which booking is possible. Label the hours of departure of these flights as $H(d,t) = \{h_1, \ldots, h_{K(d,t)}\}$, with values from 1 to 15, and where it is possible that $h_i = h_j$ when two or more flights are scheduled to leave during the same hour.[2] Using this notation, all observed booking counts (including occurrences of zero bookings) made t days before departure on flights departing on day d, can be stacked into a $K(d,t)$ vector of varying length $\boldsymbol{y}^{d,t} = (y_1^{d,t}, y_2^{d,t}, \ldots, y_{K(d,t)}^{d,t})$.[3] Its joint distribution function is also given by the copula decomposition. It is a property of the Gaussian copula that it is closed under marginalization, including for a subset of elements as here, so that the distribution function of $\boldsymbol{y}^{d,t}$ is

$$F(\boldsymbol{y}^{d,t}; \Omega) = \Phi_{K(d,t)}(y_1^{\star,d,t}, y_2^{\star,d,t}, \ldots, y_{K(d,t)}^{\star,d,t}; \Omega^{d,t})\,.$$

Here, $\Omega^{d,t} = \{\omega_{i,j}^{d,t}\}$ is a $K(d,t) \times K(d,t)$ matrix formed from $\Omega = \{\omega_{i,j}\}$, by setting element $\omega_{i,j}^{d,t} = \omega_{h_i,h_j}$ for $i = 1, \ldots, K(d,t)$ and $j = 1, \ldots, K(d,t)$. That is, $\Omega^{d,t}$ is function of Ω formed by simply 'pulling out' the relevant elements. The latent

[1]As an illustrative example, if on day d_1 flights were scheduled to depart at 07:15, 07:30, 9:30 and 10:30, then the vector would be of length $K(d_1,t) = 4$. If 30 days prior to departure the 9:30 flight was cancelled, then the vector would of length $K(d_1,t) = 3$ for bookings with $t \leq 30$. And if on the next departure day $d_1 + 1$ there are additional flights also scheduled to depart at 11:00 and 11:30, then $K(d_1+1,t) = 6$ with no cancellations.

[2]To continue the illustrative example, $H(d_1,t) = (2,2,4,5)$ for $t > 30$, $H(d_1,t) = (2,2,5)$ for $t \leq 30$ and $H(d_1+1,t) = (2,2,4,5,6,6)$.

[3]To further continue the illustrative example, if there were 3 and 6 bookings on day $t > 30$ for the flights departing at 7:15 and 9:30 on day d_1, respectively, then $\boldsymbol{y}^{d_1,t} = (3,0,6,0)$.

variables are distributed $\boldsymbol{y}^{\star,d,t} = (y_1^{\star,d,t}, y_2^{\star,d,t}, \ldots, y_{K(d,t)}^{\star,d,t}) \sim N(\mathbf{0}, \Omega^{d,t})$, constrained by the bounds as at Eqn. $\Phi^{-1}(\mathrm{Po}(y_l - 1; \lambda_l)) < y_l^\star < \Phi^{-1}(\mathrm{Po}(y_l; \lambda_l))$.

The likelihood of the proposed multivariate copula model is the product of the probability mass functions obtained from Eqn. $\Phi_{K(d,t)}(y_1^{\star,d,t}, y_2^{\star,d,t}, \ldots, y_{K(d,t)}^{\star,d,t}; \Omega^{d,t})$ over all pairs of (d, t). However, direct evaluation of each of these individual mass functions is an $O(2^{K(d,t)})$ operation, which is computationally infeasible for the values of $K(d, t)$ in our data, which are typically greater than 15. Instead, we follow Pitt et al. (2006), Danaher and Smith (2011), and Smith and Khaled (2012), and estimate the copula model using Bayesian data augmentation, which generates the constrained latent variables $\boldsymbol{y}^{\star,d,t}$ observing $\Phi^{-1}(\mathrm{Po}(y_l - 1; \lambda_l)) < y_l^\star < \Phi^{-1}(\mathrm{Po}(y_l; \lambda_l))$ using Markov chain Monte Carlo (MCMC) methods. Details are discussed next.

4.2 Copula Estimation

It is computationally infeasible to evaluate the likelihood of high dimensional copula models with discrete margins directly; for example, see the discussion in Smith and Khaled (2012). Therefore, we follow Pitt et al. (2006); Danaher and Smith (2011) and subsequent authors and estimate the copula parameters using Bayesian data augmentation. This provides estimates of the copula parameters—and associated Spearman correlations—from the Bayesian posterior distribution.

Because this is a Bayesian approach, a prior distribution for the copula parameters has to be adopted. For this, we follow Joe (2006); Daniels and Pourahmadi (2009) and parameterize Ω through its partial correlations. If $1 \leq j < i \leq 15$, these are given by $r_{i,j} = \mathrm{Corr}(y_j^\star, y_i^\star | y_{j+1}^\star, \ldots, y_{i-1}^\star)$, where the correlation is defined to be unconditional when $j = i - 1$. The set of all partials is therefore $\boldsymbol{r} = \{r_{i,j}; i = 1, \ldots, 15; j < i\}$. This parameterization is invariant with respect to the ordering of the elements of

$\boldsymbol{y}^\star$, unlike the Cholesky decomposition of Ω used in Smith and Kauermann (2011); Danaher and Smith (2011) and others. Daniels and Pourahmadi (2009) give a one-to-one transformation between Ω and $\boldsymbol{r}$, that is widely attributed to Yule.

The approach generates the latent Gaussian variables $\boldsymbol{y}^\star = \{\boldsymbol{y}^{\star,d,t}; d \in D, t = 0, \ldots, 120\}$ as part of the Markov chain Monte Carlo (MCMC) scheme below. This greatly simplifies estimation, because the posterior of $\boldsymbol{r}$ conditional on $\boldsymbol{y}^\star$ is fast to compute.

Sampling Scheme

Step 1. For $d = 1, \ldots, D$, $t = 0, \ldots, 120$, generate from $f(y_i^{\star,d,t}|\{\boldsymbol{y}^\star \backslash y_i^{\star,d,t}\}, \boldsymbol{r}, \boldsymbol{y}) = f(y_i^{\star,d,t}|\{\boldsymbol{y}^{\star,d,t} \backslash y_i^{\star,d,t}\}, \Omega^{d,t}, y^{\star,d,t})$.

Step 2. Generate from $f(\boldsymbol{r}|\boldsymbol{y}^\star)$ element-by-element using (adaptive) random walk Metropolis-Hastings.

Step 3. Compute Ω from $\boldsymbol{r}$ using Yule's one-to-one transformation.

For Step 1, note that $\boldsymbol{y}^{\star,d,t} \sim N(0, \Omega^{d,t})$, from which the mean μ and variance s^2 of the conditional distribution of the element $y_i^{\star,d,t}|\boldsymbol{y}^{\star,d,t} \backslash y_i^{\star,d,t} \sim N(\mu, s^2)$ can be computed easily. To compute the required conditional posterior, this needs to be combined with the constraint $(L_i^{d,t} < y_i^{\star,d,t} < U_i^{d,t})$, where the lower bound $L_i^{d,t} = \Phi^{-1}\left(\text{Po}(y_i^{d,t} - 1; \lambda_i^{d,t})\right)$ and the upper bound $U_i^{d,t} = \Phi^{-1}\left(\text{Po}(y_i^{d,t}; \lambda_i^{d,t})\right)$. Here, Φ is the standard normal distribution function, and $\text{Po}(\cdot; \lambda_i^{d,t})$ is the distribution function of the Poisson regression model in Section 4 with intensity value $\lambda_i^{d,t}$ for booking count $y_i^{d,t}$. (Note that we define $\text{Po}(-1; \lambda) = -\infty$ here). The conditional posterior in Step 1 is therefore a $N(\mu, s^2)$ distribution constrained to the range $(L_i^{d,t}, U_i^{d,t}]$. The bounds are computed only once, based on the fitted Poisson regression model, so that it is fast to sample each element. Moreover, the elements can be sampled in parallel because the loops in d and t are not recursive.

To implement the random walk Metropolis-Hastings (MH) in Step 2, $f(\boldsymbol{r}|\boldsymbol{y}^\star) \propto f(\boldsymbol{y}^\star|\boldsymbol{r})f(\boldsymbol{r})$, where the prior $f(\boldsymbol{r})$ is flat on the partial correlations. The augmented likelihood is

$$f(\boldsymbol{y}^\star|\boldsymbol{r}) = \prod_d \prod_t \phi_{K(d,t)}(\boldsymbol{y}^{\star,d,t}; \mathbf{0}, \Omega^{d,t}) \,.$$

By first computing Ω from $\boldsymbol{r}$ using Yule's one-to-one transformation, the matrices $\Omega^{d,t}$ above can be formed by simply extracting their elements from Ω. The density is then evaluated directly, which requires the Cholesky factorization of each matrix $\Omega^{d,t}$. When programmed in a low level language (Fortran 90) we found this is practical to implement on regular PCs with the sample sizes examined here. Moreover, the products can be readily computed in parallel, greatly speeding the evaluation. Note that all computations are undertaken on the logarithmic scale for numerical stability, as is usually the case when implementing a MH step. In general, we run our sampling scheme for a burnin of 40,000 iterates, and a collect a further 20,000 iterates from which to compute posterior inference, which takes around 4 hours on a standard desktop for our dataset.

4.3 Estimated Dependence

As in Section 3, we fit the model separately for different departure day types. We also further segment by the number of days prior to departure when the booking was made, and by the booking day type. To measure the overall level of dependence we compute the posterior estimates of the Spearman pairwise correlations between $y_i^\star$ and $y_j^\star$, which is $\rho_{i,j}^s = \frac{6}{\pi} \arcsin \omega_{i,j}$ for a Gaussian copula parameter matrix $\Omega = \{\omega_{i,j}\}$. Figure A.4 plots the posterior mean of the matrix of Spearman pairwise correlations $R = \{\rho_{i,j}^s\}$ for bookings made on weekdays for flights departing Thursdays. The panels give estimates for bookings made between (a) $2 \leq t \leq 30$, (b) $30 < t \leq 60$ and (c) $t > 60$ days prior to departure. Blank cells show where the 99% posterior probability intervals for $\rho^s_{i,j}$ contain 0. For bookings made in the month prior to departure (panel (a)), there is positive dependence throughout. This is likely due to the omission of factors that drive demand for all flights at a daily level. A similar feature can be seen with bookings made between long before departure in panels (b,c), but mostly for flights that depart in the evening. In either case, the level of dependence is only mild, suggesting the proposed Poisson model accounts for the vast majority of dependence between bookings for flights that depart on the same day. While not reported here, very similar results were found for other segmentations of the bookings data.

5 Penalized Maximum Likelihood Estimation

For simplicity of notation we write $\pi_{k,i,t}$ instead of $\pi_k(t)$ and define $\boldsymbol{\theta}_k^{(\pi)} = (\boldsymbol{\beta}_k^{(\pi)}, \boldsymbol{\gamma}_{0,k}^{(\pi)}, \boldsymbol{\gamma}_{1,k}^{(\pi)})$ as corresponding subvector of $\boldsymbol{\theta}$. The corresponding model design matrix for the i-th flight at t days to departure is denoted as

$$\boldsymbol{w}_{i,t}^{(\pi)} = \Big(\mathcal{I}(BDAY_i = j), j = 1, ..., 7; \boldsymbol{w}_0^{(\pi)}(t), \boldsymbol{w}_1^{(\pi)}(DTIME_i)\Big)$$

where $\boldsymbol{w}_0^{(\pi)}(t)$ and $\boldsymbol{w}_1^{(\pi)}(DTIME_i)$ are B-spline basis functions in time and departure time, see also Appendix B. Analogously we define

$$\boldsymbol{w}_{i,t}^{(\lambda)} = \Big(\mathcal{I}(BDAY_i = j), j = 1, ..., 7; \boldsymbol{w}_0^{(\lambda)}(t), \boldsymbol{w}_1(\lambda)(DTIME_i)\Big)$$

and $\boldsymbol{\theta}^{(\lambda)} = (\boldsymbol{\beta}^{(\lambda)}, \boldsymbol{\gamma}_0^{(\lambda)}, \boldsymbol{\gamma}_1^{(\lambda)})^T$ to be the design matrix and corresponding parameter vector for modelling λ_{BL}. Finally for the group specific part δ_k we define the matrix as $\boldsymbol{v}_{i,t} = (\text{PRICE}_{i,t})$ or $\boldsymbol{v}_{i,t} = (\text{PRICE}_{i,t}, \widehat{\xi}_{i,t})$ depend on whether we fit the model without or with instrumental variable where $\text{PRICE}_{i,t}$ is the price for flight i at t days to departure and $\hat{\xi}_{i,t}$ the fitted residual of the OLS estimation of Eqn. (3.7). The matching vector of parameters is $\boldsymbol{\alpha}_k$. Then the first partial derivatives, defining the gradients, are:

5.1 Derivatives

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k^{(\pi)}} = \sum_i \sum_t \boldsymbol{w}_{i,t}^{(\pi)^T} \Big(\frac{Y_{i,t}}{\lambda_{i,t}} - 1\Big) \lambda_{0,i,t} \Big(\pi_{k,i,t}(1-\pi_{k,i,t})(\delta_{k,i,t} - \delta_{K,i,t})\Big)$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(\lambda)}} = \sum_i \sum_t \boldsymbol{w}_{i,t}^{(\lambda)^T} (Y_{i,t} - \lambda_{i,t})$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha_k} = \begin{cases} \sum_i \sum_t \boldsymbol{v}_{i,t}^T \Big(\frac{Y_{i,t}}{\lambda_{i,t}} - 1\Big) & \text{if } k = 1 \\ -c_k \Bigg(\sum_i \sum_t \boldsymbol{v}_{i,t}^T \Big(\frac{Y_{i,t}}{\lambda_{i,t}} - 1\Big) \lambda_{0,i,t} \Big(\sum_k \pi_{k,i,t} \lambda_{k,i,t} \Big) \Bigg) & \text{if } k > 1 \end{cases}$$

where $\lambda_{0,i,t} = \exp(\boldsymbol{w}_{i,t}^{(\lambda)} \boldsymbol{\theta}^{(\lambda)})$, $\delta_{k,i,t} = \exp(\boldsymbol{v}_{i,t} \boldsymbol{\gamma}_k)$ and $c_k = \exp(\alpha_k)$. The Fisher information results as

$$\mathbb{E}\Bigg(- \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{k_1}^{(\pi)} \partial \boldsymbol{\theta}_{k_2}^{(\pi)^T}} \Bigg) = \begin{cases} \sum_i \sum_t \boldsymbol{w}_{i,t}^{(\pi)^T} \lambda_{0,i,t}^2 \Bigg(\frac{\pi_{k,i,t}^2 \left(1-\pi_{k,i,t}\right)^2 \left(\lambda_{k,i,t} - \lambda_{K,i,t}\right)^2}{\lambda_{i,t}} \Bigg) \boldsymbol{w}_{i,t}^{(\pi)} & \text{if } k = k_1 = k_2 \\ \sum_i \sum_t \boldsymbol{w}_{i,t}^{(\pi)^T} \lambda_{0,i,t}^2 \Big(\prod_{k \in \{k_1, k_2\}} \pi_{k,i,t} (1 - \pi_{k,i,t}) \Big) B_{k,i,t} \boldsymbol{w}_{i,t}^{(\pi)} & \text{if } k_1 \neq k_2 \end{cases}$$

$$\mathbb{E}\Bigg(- \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(\lambda)} \partial \boldsymbol{\theta}^{(\lambda)^T}} \Bigg) = \boldsymbol{w}_{i,t}^{(\lambda)^T} \lambda_{,i,t} \boldsymbol{w}_{i,t}^{(\lambda)}$$

$$\mathbb{E}\Bigg(- \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}_{k_1} \partial \boldsymbol{\alpha}_{k_2}^T} \Bigg) = \begin{cases} \sum_i \sum_t \boldsymbol{v}_{i,t}^T \lambda_{it} \boldsymbol{v}_{i,t} & \text{if } k = k_1 = k_2 = 1 \\ c_k \Bigg(\sum_i \sum_t \boldsymbol{v}_{i,t}^T \lambda_{0,i,t}^3 \frac{A_{k,i,t}^2}{\lambda_{it}} \boldsymbol{v}_{i,t} \Bigg) c_k & \text{if } k = k_1 = k_2 > 1 \\ -c_{k_2} \Big(\sum_i \sum_t \boldsymbol{v}_{i,t}^T \lambda_{0,i,t} A_{k_2,i,t} \boldsymbol{v}_{i,t} \Big) & \text{if } k_1 = 1, k_2 > 1 \\ c_{k_1} \Bigg(\sum_i \sum_t \boldsymbol{v}_{i,t}^T \lambda_{0,i,t}^3 \frac{A_{k_1,i,t} A_{k_2,i,t}}{\lambda_{i,t}} \boldsymbol{v}_{i,t} \Bigg) c_{k_2} & \text{if } k_1, k_2 > 1, \quad k_1 \neq k_2 \end{cases}$$

where $A_{k,i,t} = \Bigg(\sum_{j=k}^{K} \pi_{j,i,t} \lambda_{j,i,t} \Bigg)$ and $B_{k,i,t} = \frac{\lambda_{k,i,t} - \lambda_{K,i,t}}{\lambda_{i,t}}$

$$\mathbb{E}\left(-\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{k_1}^{(\pi)} \partial \boldsymbol{\alpha}_{k_2}{}^T}\right) = \begin{cases} -\sum_i \sum_t \boldsymbol{w}_{i,t}^{(\pi)^T} \lambda_{0,i,t} \pi_{k_1,i,t} (1-\pi_{k_1,i,t}) B_{k_1,i,t} \boldsymbol{v}_{i,t} & \text{if } k_1 \geq 1, k_2 = 1 \\ -c_{k_2} \left(\sum_i \sum_t \boldsymbol{w}_{i,t}^{(\pi)^T} \lambda_{0,i,t}^2 \pi_{k_1,i,t} (1-\pi_{k_1,i,t}) A_{k_2,i,t} B_{k_1,i,t} \boldsymbol{v}_{i,t} \right) & \text{if } k_1 \geq 1, k_2 > 1 \end{cases}$$

$$\mathbb{E}\left(-\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{k}^{(\pi)} \partial \boldsymbol{\theta}^{(\lambda)^T}}\right) = \sum_i \sum_t \boldsymbol{w}_{i,t}^{(\pi)^T} \lambda_{\mathrm{BL}} (\lambda_{k,i,t} - \lambda_{K,i,t}) \pi_{k,i,t} (1-\pi_{k,i,t}) \boldsymbol{w}_{i,t}^{(\lambda)}$$

$$\mathbb{E}\left(-\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{k}^{(\lambda)} \partial \boldsymbol{\alpha}_{k}^T}\right) = \begin{cases} \sum_i \sum_t \boldsymbol{w}_{i,t}^{(\lambda)^T} \lambda_{i,t} \boldsymbol{v}_{i,t} & \text{if } k = 1 \\ -c_k \left(\sum_i \sum_t \boldsymbol{w}_{i,t}^{(\lambda)^T} \lambda_{0,i,t} A_{k,i,t} \boldsymbol{v}_{i,t} \right) & \text{if } k > 1 \end{cases}$$

As (3.2) and (3.4) are typically not identifiable if the B-splines basis is used a mean centering constraint, see e.g. Wood (2017), is applied to each smooth component. For instance centering the component $\boldsymbol{w}_0^{(\pi)}(t)$ is achieved by finding the matrix $\boldsymbol{Z}_0$ which solves $\mathbf{1}^T \boldsymbol{w}_0^{(\pi)}(t) \boldsymbol{Z}_0 = \mathbf{0}$ where $\boldsymbol{Z}_0$ has one column less then the original design-matrix $\boldsymbol{w}_0^{(\pi)}(t)$. By the use of the re-parameterized parameter-vector $\boldsymbol{\gamma}_{0,k,c}^{(\pi)} = \boldsymbol{Z}_0 \boldsymbol{\gamma}_{0,k}^{(\pi)}$ for estimation, the centering constraint is automatically satisfied.

5.2 Penalization Setting

Based on the ideas of Eilers and Marx (1996) and Ruppert et al. (2003). We impose a penalty on the coefficients relating to the functional effects $s_{0k}^{(\pi)}(\mathrm{t})$, $s_{1k}^{(\pi)}(\mathrm{DTIME})$,$s_0^{(\lambda)}(\mathrm{t})$ and $s_1^{(\lambda)}(\mathrm{DTIME})$, respectively. We make use of linear B-splines and penalize neighboring coefficients. To be specific we set $\boldsymbol{w}_0^{(\pi)}(\mathrm{t})$ as linear B-spline bases with 12 knots located at equidistantly between -11 and 133. We therefore penalize first order differences of the components of $\gamma_{0k}^{(\pi)}$, i.e., $\gamma_{0,k,l}^{(\pi)} - \gamma_{0,k,l-1}^{(\pi)}, l = 10, \ldots, 2$. Analogously we specify the remaining spline base matrices. The penalties can be written as quadratic form leading to the penalized likelihood $\ell_p(,)$

$$\ell_p(\boldsymbol{\theta}, \boldsymbol{\rho}) = \ell(\boldsymbol{\theta}) + \sum_{j=0}^{1} \sum_{k=1}^{K-1} \rho_{jk}^{(\pi)} \boldsymbol{\gamma}_{jk}^{(\pi)^T} D_{jk}^{(\pi)} \boldsymbol{\gamma}_{jk}^{(\pi)} + \sum_{j=0}^{1} \rho_j^{(\lambda)} \boldsymbol{\gamma}_j^{(\lambda)^T} D_j^{(\lambda)} \boldsymbol{\gamma}_j^{(\lambda)}$$

where $\boldsymbol{\rho} = (\rho_0^{(\lambda)}, \rho_1^{(\lambda)}, \rho_{0k}^{(\lambda)}, \rho_{1k}^{(\lambda)}, k = 1, ..., K)$ are the penalty parameters to be specified later. Apparently, if $\boldsymbol{\rho} = 0$ one obtains unpenalized estimation. The smoothing matrices D result from taking differences of neighboring coefficients and exactly follows the convention of Eilers and Marx (1996). This means, for instance, that the difference of spline coefficients is penalized so that neighbouring spline coefficients are forced to be of similar size.

The penalty parameters need to be selected, data driven and we here use the Bayesian Information Criterion (BIC) defined through

$$BIC(\boldsymbol{\rho}) = -2\ell(\hat{\boldsymbol{\theta}}) + \log(n)\mathrm{df}(\boldsymbol{\rho})$$

where $\mathrm{df}(\boldsymbol{\rho})$ is the degree of the model and n is the number of observations ($\approx$ number of flights multiplied by the number of considered days to departure). The degree of the model can be approximated through Fisher matrices as follows. Let $F(\boldsymbol{\theta}, \boldsymbol{\rho})$ denote the penalized Fisher matrix, i.e.

$$F(\boldsymbol{\theta}, \boldsymbol{\rho}) = E\left(-\frac{\partial \ell_p(\boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right).$$

Then, the degree of the model is approximated through

$$\mathrm{df}(\boldsymbol{\rho}) = \mathrm{trace}\{F^{-1}(\hat{\boldsymbol{\theta}}, \boldsymbol{\rho}) F(\hat{\boldsymbol{\theta}}, \boldsymbol{\rho} = \mathbf{0})\}$$

where $\hat{\boldsymbol{\theta}}$ is the penalized parameter estimate. For a justification of (A.6) see Ruppert et al. (2003) or Krivobokova and Kauermann (2007).

We maximize (5.2) and apply for simplicity the same degree of smoothing for t and DTIME. That is we set $\rho_{jk}^{(\pi)} = \rho_j^{(\lambda)}\ \forall\ j = 0, 1$. If $K = 2$ within the optimization procedure the control parameters $\rho_{01}^{(\pi)} = \rho_0^{(\lambda)}$ and $\rho_{11}^{(\pi)} = \rho_1^{(\lambda)}$ will be fixed to some value. These values are selected based on a grid search by minimizing the BIC.

6 Bootstrapping for Mixture Models

We compute bootstrap confidence intervals using the 'leave out one individual' approach of Rice and Silverman (1991). Re-sampling is undertaken on the flight level to accommodate dependence between bookings on the same flight, and consistent with the likelihood at Eqn. (3.6). For each flight i, booking counts and associated covariates are re-sampled (with replacement) for the entire window of booking days between t_i^{close} and t_i^{open}.

To control for label switching we choose to order the segment specific price coefficients α_k in a monotone sequence. The label switching problem occurs for such random samples from its population whenever at least two group labels $\delta_k, k = 1, \ldots, K$ from Eqn. (3.1) change their positions. If the superscript l within δ_k^l denotes the group label for a random sample, than there exists at least two group indices k for which $\delta_k^1 \neq \delta_k^2$ if a single label switch between two groups occurs.

To avoid label switching and its negative impact on confidence intervals that result from re-sampling techniques such as boostraping, we propose a frequentist control approach.

To identify the segment labels $\delta_k = \exp(\alpha_{1,k}\text{PRICE} + \alpha_{2,k}\hat{\xi})$ in 3.1 we impose ordering constraints on the elements of the coefficient vector $\boldsymbol{\alpha}_j^T = (\alpha_{j,1}, \ldots, \alpha_{j,K})$. The identification of the group labels is finally achieved by constraints the elements of $\boldsymbol{\alpha}_j$ such that $\alpha_{j,k} < \alpha_{j,k+1}$ or that $\alpha_{j,k} > \alpha_{j,k+1}$. Allowing for every possible ordering $(<, >)$ between two neighboring coefficients $\alpha_{j,k}, \alpha_{j,k+1}$, a total of K^J possible ordering combinations result. For Eqn. 3.1, we have $J = 2$ and if $K = 2$, the number of possible ordering combinations is 4. This number reduces to $K^{(J-1)}$ if it is acknowledged that the same ordering of $\alpha_{j,k} < \alpha_{j,k+1}$ is achieved by $\alpha_{j,k} > \alpha_{j,k+1} \forall j = 1, \dots, J, k = 1, \dots, K$ if the grouping index k no longer runs from the lowest to the highest index but rather from the highest to the lowest, i.e., $\alpha_{j,k} > \alpha_{j,k+1}$ transforms into $\alpha_{j,k+1} > \alpha_{j,k}$. Abbreviating $\alpha_{j,k} < \alpha_{j,k+1}$ by decr_j and $\alpha_{j,k} > \alpha_{j,k+1}$ by incr_j, for Eqn. (3.1) , with $J = 2$ and $K = 2$, every possible ordering constraints belongs to the set $\{\{\mathrm{decr}_1, \mathrm{decr}_2\}, \{\mathrm{decr}_1, \mathrm{incr}_2\}\}$. For a fixed value of K, a separate estimation of (3.6) for each ordering constraint is performed and the model with the lowest BIC values among all candidate models is finally chosen and the corresponding ordering constraint is used to derive bootstrap confidence bands. As the number of possible ordering constraints gets large for small values of K, we use the ordering constraints of the $K = 2$ case for $K > 2$. A consequence of this restriction is that only monotone decreasing or increasing ordering constraints between group parameters are allowed, even though the true ordering between group parameters $\alpha_{j,k}$ is possibly different. Therefore, applying the minimum BIC rule to models with $K > 2$ to select the optimal number of groups only results in an upper bound and is therefore not exact. Exemplary let $K = 3$ be the true group size and $\alpha^*_{j,1} < \alpha^*_{j,k+1} > \alpha^*_{j,3}$ the optimal ordering constraint. As we only allow for $\alpha_{j,1} < \alpha_{j,2} < \alpha_{j,3}$ or $\alpha_{j,1} > \alpha_{j2} > \alpha_{j,3}$, the BIC rule selects a model with $K = 4$ groups as a convex combination of the parameters with ordering constraint $\alpha_{j,1} < \alpha_{j,2} < \alpha_{j,3} < \alpha_{j,4}$ is able to express the relation of $\alpha^*_{j,1} < \alpha^*_{j,2} > \alpha^*_{j,3}$ by setting $\alpha^*_{j,1} = \alpha_{j,1}, \alpha^*_{j,2} = \alpha_{j,3}$, and $\alpha^*_{j,3} = \pi_2(.)\alpha_{j,2} + \pi_4(.)\alpha_{j,4}$.

7 Two-step Estimator by residual inclusion

We consider the two-step estimator of Marra and Radice (2011) to account for price-endogeneity. As the technical discussions of Marra and Radice (2011) concerns the omitted variable bias problem as a characteristic of endogeneity, some minor adjustments are necessary to provide a similar statement for the case of simultaneity.

Given that the number of arriving passengers are specified by $Y(t) = \lambda(t) + u_\lambda$, the systematic component $\lambda(t)$ characterizes through the segment-specific Eqn. $\log(\delta_k)$ (3.1) how demand depends on PRICE as:

$$Y(t) = \lambda_{\mathrm{BL}}(t) \left(\sum_{k=1}^{K} \pi_k(t) \underbrace{\exp\left(\alpha_{1,k}\mathrm{PRICE}\right)}_{\delta_k} \right) + u_\lambda$$

If PRICE is endogenous the assumption of $\mathbb{E}\left(u_\lambda \mid \mathrm{PRICE}, \mathrm{BDAY}, \mathrm{DTIME}, t\right) = 0$ is violated and the estimation of the demand-equation results in biased estimates. For $\eta = \eta\big(\mathrm{IV,BDAY,t,DTIME}\big)$ we plug the expression $\mathrm{PRICE} = \exp\left(\eta + \sigma^2/2\right) + \xi$ into the demand-equation which results in

$$Y(t) = \lambda_{\mathrm{BL}}(t) \left(\sum_{k=1}^{K} \pi_k(t) \exp\left(\alpha_{1,k} \Big\{ \exp\left(\eta + \sigma^2/2\right) + \xi \Big\} \right) \right) + u_\lambda$$

As the unobservable error ξ enters the segment-specific equations, biased estimates result if $\mathbb{E}\left(u_\lambda \mid \xi\right) \neq 0$. For the additive separation of ξ into two parts

$$\xi = \xi_1 + \xi_2.$$

we assume that $\mathbb{E}\left(u_\lambda \mid \ \xi_1\right) = 0$ but $\mathbb{E}\left(u_\lambda \mid \xi_2\right) \neq 0$. If ξ_2 would be observable we could include this variable to (3.7) as an additional regressor. Thus, the new predictor changes to

$$\eta_{new} = \theta_0 + \theta_1 IV + \sum_{j=2}^{7} \mathcal{I}(\mathrm{BDAY} = j)\theta_j + f_0(t) + f_1(\mathrm{DTIME}) + \theta_8 \xi_2$$

Therefore, the update on the PRICE-equation is

$$\text{PRICE} = \exp\left(\eta_{new} + \frac{\sigma^2_{new}}{2}\right) + v$$

Taking the Taylor approximation of $\exp\left(\eta_{new} + \frac{\sigma^2_{new}}{2}\right)$ around $\theta_8\xi_2 = 0$ results in

$$\text{PRICE} = \exp(\eta) + \underbrace{\frac{\partial \exp(\eta_{new})}{\partial \theta_8 \xi_2}\theta_8\xi_2 + v}_{:=\zeta}$$

Thus estimation of the reduced form of price with Eqn. (3.7) with instrument IV gives an estimate of ζ that contains information about the unobservable variable ξ_2. Therefore, the inclusion of the estimate $\widehat{\xi}$ within the segment-specific equations controls for the endogeneity of price.

8 Source-Code and Data-Files

The associated R- and Fortran-code for the estimation algorithm used, as well as the data for flights departing on a Thursday, can be downloaded at https://github.com/JFMeyer2k/SMIJ.

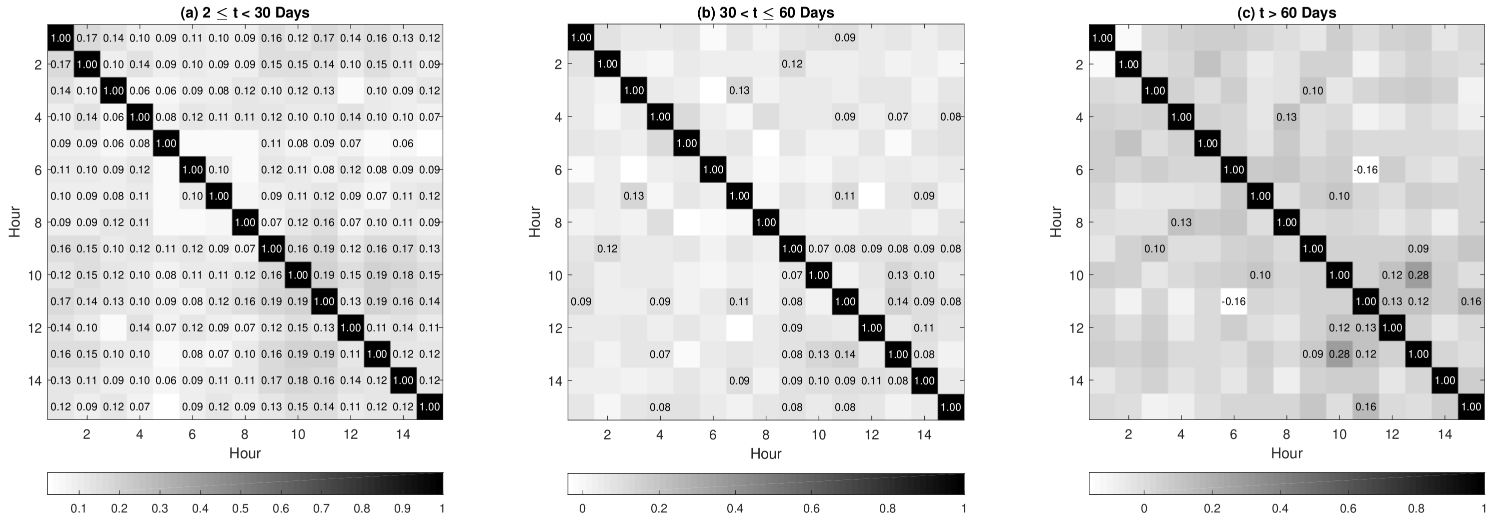

Figure A.4: Estimates of pairwise Spearman correlations from the copula model fitted to bookings made on weekdays for flights that depart on Thursdays. Each panel plots the posterior mean of $R = \{\rho^S_{i,j}\}$, arranged as a matrix for hourly indices $i = 1, \ldots, 15$ and $j = 1 \ldots, 15$. Each cell contains the posterior mean of $\rho^S_{i,j}$, except when the 99% posterior probability interval for$\rho^S_{i,j}$ includes 0, in which case the cell is left blank. Results are given for bookings further segmented by (a) 2–30, (b) 31–60, (c) 60-120 days prior to departure.

References

Allenby, G. M. and Rossi, P. E. (1998). Marketing models of consumer heterogeneity. *Journal of Econometrics*, **89**(1-2), 57–78.

Balaiyan, K., Amit, R. K., Malik, A. K., Luo, X., and A., A. (2019). Joint forecasting for airline pricing and revenue management. *J. Revenue Pricing Management*, **18** (6), 465–482. doi: 10.1057/s41272-019-00188-. URL https://ideas.repec.org/a/pal/jorapm/v18y2019i6d10.1057_s41272-019-00188-4.html.

Besbes, O. and Zeevi, A. (2015). On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. *Management Sci.*, **61**(4), 723–739. doi: 10.1287/mnsc.2014.2031. Available at http://dx.doi.org/10.1287/mnsc.2014.2031.

Bhat, C. R. and Eluru, N. (2009). A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. *Transportation Res. B*, **43** (7), 749–765.

Brons, M., Pels, R., Nijkamp, P., and Rietveld, P. (2002). Price elasticities of demand for passenger air travel: a meta-analysis. *J. Air Transp. Management*, **8** (3), 165–175. doi: http://dx.doi.org/10.1016/S0969-6997(01)00050-3. Available at http://www.sciencedirect.com/science/article/pii/S0969699701000503.

Dai, J., Ding, W., Kleywegt, A., Wang, X., and Zhang, Y. (2014). Choice based revenue management for parallel flights. *Available at SSRN 2404193*. Available at http://ssrn.com/abstract=2404193.

Danaher, P. and Smith, M. (2011). Modeling multivariate distributions using copulas: Applications in marketing. *Marketing Sci.*, **30**(1), 4–21. Available at http://EconPapers.repec.org/RePEc:inm:ormksc:v:30:y:2011:i:1:p:4-21.

Daniels, M. J. and Pourahmadi, M. (2009). Modeling Covariance Matrices via Partial Autocorrelations. *Journal of Multivariate Analysis*, **100**(10), 2352–2363.

Davidson, R. and MacKinnon, J. (1993). *Estimation and Inference in Econometrics.* Oxford University Press, 1st edition.

Davidson, R. and MacKinnon, J. (1999). *Econometric Theory and Methods.* Oxford University Press, 1st edition.

Diaconescu, E. (2008). The use of NARX neural networks to predict chaotic time series. *Wseas Transactions on computer research*, **3**(3), 182–191.

Dissmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, **59**, 52–69.

Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties. *Stat. Sci.*, **11**, 89–121.

Eluru, N., Paleti, R., Pendyala, R., and Bhat, C. (2010). Modeling injury severity of multiple occupants of vehicles: Copula-based multivariate approach. *Transportation Research Record: Journal of the Transportation Research Board*, **2165**(1), 1–11.

Escobari, D. (2017). Airport, airline and departure time choice and substitution patterns: An empirical analysis. *Transportation Res. A*, **103**, 198–210.

Feldman, J. and Topaloglu, H. (2015). Bounding optimal expected revenues for assortment optimization under mixtures of multinomial logits. *Prod. Oper. Management*, **24**(10), 1598–1620. Available at http://EconPapers.repec.org/RePEc:bla:popmgt:v:24:y:2015:i:10:p:1598-1620.

Fiig, T., Hardling, R., Pölt, S., and Hopperstad, C. (2014). Demand forecasting and measuring forecast accuracy. *J. Revenue Pricing Management*, **13**(6), 413–439. Available at http://dx.doi.org/10.1057/rpm.2014.29.

Granados, N., Kauffman, R., Lai, H., and Lin, H.-C. (2012). À la carte pricing and price elasticity of demand in air travel. *Decis. Support Syst.*, **53**(2), 381–394.

Guevara, C. A. (2018). Overidentification tests for the exogeneity of instruments in discrete choice models. *Transportation Res. B*, **114**(C), 241–253.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models.* London: Chapman & Hall.

Hetrakul, P. and Cirillo, C. (2014). A latent class choice based model system for railway optimal pricing and seat allocation. *Transportation Res. E*, **61**(C), 68–83.

IATA. IATA Industry Facts & Statistics.

IATA. IATA Industry Facts & Statistics.

Joe, H. (2006). Generating Random Correlation Matrices based on Partial Correlations. *Journal of Multivariate Analysis*, **97**(10), 2177–2189.

Jordan, M. I. and Jacobs, R. A. (1994). Hierarchical mixtures of experts and the em algorithm. *Neural computation*, **6**(2), 181–214.

Jung, J. M. and Fujii, E. T. (1979). The prize elasticity of demand for air travel. *J. Transp., Econ., and Policy*, **10**, 257–262.

Karlis, D. and Xekalaki, E. (2005). Mixed poisson distributions. *Int. Stat. Rev.*, **73** (1), 35–58.

Kauermann, G., Krivobokova, T., and Fahrmeir, T. (2009). Some asymptotic results on generalized penalized spline smoothing. *J. of the Royal Statistical Society, Series B*, **71**(2), 487–503.

Ke, J., Zheng, H., Yang, H., and Chen, X. M. (2017). Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Res. C*, **85**, 591–608.

Kremers, H., Nijkamp, P., and Rietveld, P. (2002). A meta-analysis of price elasticities of transport demand in a general equilibrium framework. *Economic Modelling*, **19**, 463–485.

Krivobokova, T. and Kauermann, G. (2007). A note on penalized spline smoothing with correlated errors. *J. of the American Statistical Association*, **102**(480), 1328–1337. doi: 10.1198/016214507000000978.

Li, J., Granados, N., and Netessine, S. (2014). Are consumers strategic? structural estimation from the air-travel industry. *Management Sci.*, **60**(9), 2114–2137.

Lin, L., He, Z., and Peeta, S. (2018). Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. *Transportation Res. C*, **97**, 258–276.

Lo, W., Wan, Y., and Zhang, A. (2015). Empirical estimation of price and income elasticities of air cargo demand: The case of hong kong. *Transportation Res. A*, **78**, 309–324. ISSN 0965-8564. doi: http://dx.doi.org/10.1016/j.tra.2015.05.014. Available at http://www.sciencedirect.com/science/article/pii/S0965856415001421.

Lurkin, V., Garrow, L. A., Higgins, M. J., Newman, J. P., and Schyns, M. (2017).

Accounting for price endogeneity in airline itinerary choice models: An application to continental us markets. *Transportation Res. A*, **100**, 228–246.

Marra, G. and Radice, R. (2011). A flexible instrumental variable approach. *Stat. Model.*, **11**(6), 581–603.

Marra, G. and Wood, S. N. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian J. Statist.*, **39**(1), 53–74. Available at https://EconPapers.repec.org/RePEc:bla:scjsta:v:39:y:2012:i:1:p:53-74.

Martinez-Garcia, E. and Royo-Vela, M. (2010). Segmentation of low-cost flights users at secondary airports. *J. Air Transp. Management*, **16**, 234–237.

McGill, J. and Van Ryzin, G. (1999). Revenue management: Research overview and prospects. *Transportation Sci.*, **33**(2), 233–256.

Mumbower, S., Garrow, L., and Higgins, M. (2014). Estimating flight-level price elasticities using online airline data: A first step toward integrating pricing, demand, and revenue optimization. *Transportation Res. A*, **66**, 196 – 212.

Oum, T. H., Waters, W. G., and Yong, J. S. (1992). Concepts of price elasticities of transport demand and recent empirical estimates: An interpretative survey. *J. Transp., Econ., and Policy*, **26**, 139–154.

Peterson, O. and Thankom, A. (2020). Spillover of covid-19: Impact on the global economy. *MPRA Paper*, (99317). URL https://ideas.repec.org/p/pra/mprapa/99317.html.

Petrin, A. and Train, K. (2010). A control function approach to endogeneity in consumer choice models. *J. Marketing Res.*, **47**, 3–13.

Pitt, M., Chan, D., and Kohn, R. (2006). Efficient bayesian inference for gaussian copula regression models. *Biometrika*, **93**(3), 537–554. Available at http://www.jstor.org/stable/20441306.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. of the Royal Statistical Society, Series B*, **53**(1), 233–243.

Ruppert, D., Wand, M., and Carroll, R. (2003). *Semiparametric Regression.* Cambridge University Press.

Ruppert, D., Wand, M., and Carroll, R. (2009). Semiparametric regression during 2003 - 2007. *Electron. J. Stat.*, **3**, 1193 – 1256.

Sfeir, G., Abou-Zeid, M., Rodrigues, F., Camara Pereira, F., and Kaysi, I. (2021). Latent class choice model with a flexible class membership component: A mixture model approach. *Journal of Choice Modelling*, **41**, 100320. ISSN 1755-5345. doi: https://doi.org/10.1016/j.jocm.2021.100320. URL https://www.sciencedirect.com/science/article/pii/S1755534521000531.

Smith, M. S. and Kauermann, G. (2011). Bicycle commuting in melbourne during the 2000s energy crisis: A semiparametric analysis of intraday volumes. *Transportation Res. B*, **45**(10), 1846–1862.

Smith, M. and Khaled, M. (2012). Estimation of copula models with discrete margins via bayesian data augmentation. *J. of the American Statistical Association*, **107** (497), 290–303.

Song, P.-K. (2000). Multivariate dispersion models generated from gaussian copula. *Scandinavian J. Statist.*, **27**(2), 305–320. Available at http://EconPapers.repec.org/RePEc:bla:scjsta:v:27:y:2000:i:2:p:305-320.

Talluri, K. and van Ryzin, G. (2004). Revenue management under a general discrete choice model of consumer behavior. *Management Sci.*, **50**(1), 15–33.

Talluri, K. and van Ryzin, G. (2005). *The Theory and Practice of Revenue Management*. International Series in Operations Research & Management Science. Springer US.

Teichert, T., Shehu, E., and von Wartburg, I. (2008). Customer segmentation revisited: The case of the airline industry. *Transportation Res. A*, **42**(1), 227–242.

Tutz, G. (2012). *Regression for Categorical Data*. Cambridge University Press.

Van Ryzin, G. and Vulcano, G. (2014). A market discovery algorithm to estimate a general class of nonparametric choice models. *Management Sci.*, **61**(2), 281–300.

Vij, A. and Walker, J. L. (2014). Preference endogeneity in discrete choice models. *Transportation Res. B*, **64**, 90–105.

Vulcano, G., van Ryzin, G., and Chaar, W. (2010). Choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing Service Oper. Management*, **12**(3), 371–392.

Vulcano, G., van Ryzin, G., and Ratliff, R. (2012). Estimating primary demand for substitutable products from sales transaction data. *Oper. Res.*, **60**(2), 313–334.

Wedel, M. and Kamakura, W. A. (2012). *Market segmentation: Conceptual and methodological foundations*, volume 8. Springer Science & Business Media.

Wen, C.-H. and Lai, S.-C. (2010). Latent class models of international air carrier choice. *Transportation Res. E*, **46**(2), 211–221.

Wen, C.-H. and Chen, P.-H. (2017). Passenger booking timing for low-cost airlines: A continuous logit approach. *J. Air Transp. Management*, **64**, 91–99.

Whittaker, T. A. and Miller, J. E. (2021). Exploring the enumeration accuracy of cross-validation indices in latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, **28**(3), 376–390.

Wood, S. (2017). *Generalized Additive Models: An Introduction with R*. Chapman & Hall, CRC texts in statistical science. CRC Press.

Wooldridge, J. (2002). *Introductory Econometrics: A Modern Approach*. South-Western, 2nd edition.

Xu, C., Ji, J., and Liu, P. (2018). The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transportation Res. C*, **95**, 47–60.