The Pricing Behaviour of Firms in the On-line Accommodation Market: Evidence from a Metropolitan City

Andrea Guizzardi and Flavio Maria Emanuele Pons

Abstract The widespread diffusion of on-line travel agencies has opened the possibility, for hoteliers, to update continously quality and prices offered along the advance booking. We study firms' pricing behaviour in a business-oriented environment considering time series of daily best available rates for 107 hotels in Milan, over a period of 9 months, from 0 to 28 days of advance booking. Throught a panel-VAR approach we assess if the typical planning of the price trajectory, including dummies for holidays and fairs as covariates. Results suggest that strategies put into effect by firms reflect some of the basic principles of the online revenue management. Price trajectories are planned considering both firms expectations on the prices they hope to charge last-minute, and their need to guarantee price stability along the advance booking. Fairs and holidays show a different impact on price dynamics. While the response caused by an "holiday shock" tends to be flat, room rates during fairs raise in the immediate future, then accommodate to the equilibrium in about three days.

Key words: RevPOR, RevPAR, Dynamic Pricing, panel-VAR

1 Introduction

In recent years, rapid technological progress and the proliferation of on-line booking platforms have deeply modified the behaviour of touristic firms, in particular concerning pricing and the management of the occupation rate. The opportunity to

AndreaGuizzardi
Department of Statistics, University of Bologna, Via delle Belle Arti 41, Bologna, e-mail: andrea.guizzardi@unibo.it

Flavio Maria Emanuele Pons
Department of Statistics, University of Bologna, Via delle Belle Arti 41, Bologna, e-mail: flaviomaria.pons2@unibo.it

update quality and prices of the room offered on-line in real time has boosted the development of new methods and algorithms to perform an effective revenue management. Moreover, the great availability of free data deriving from on-line travel agencies increases the transparency of the market and the possibility to study firm competition.

Considering the industry characteristics (perishable inventory, short-term constrained capacity, high fixed costs respect to variable costs) managers are motivated to adjust prices in real time in order to sell all the rooms out by the target day emphasizing maximization of daily revenues ([Wang and Brennan, 2014]). However, hotels operate a segmentation of the room market, considering that there exists a negative correlation between the ability to purchase and the advance between booking and arrival date. In fact, clients with higher spending possibility, in general business travellers, tend to make a reservation on the target day or with a short notice of one or two days ([Guo et al., 2013]).

This situation induces a trade-off between the strong incentive to sell all the rooms by the target day (minimizing strategically - unsold capacity) and the higher profitability of the rooms sold with a short advance booking (maximizing tactically - average Revenue Per Occupied Room (RevPOR)). In the on-line market, the advance booking enables firms to mix strategic and tactic pricing, in order to maximize Revenue Per Available Room (RevPAR). This operation is not straigtforward, particularly in a location, such as Milan, where the low elasticity to price of business customers concurs in differentiating the economic effect of tactic and strategy in revenue management.

Previous researches (e.g. [Abrate and Viglia, 2016]) have exploited the full potential of a panel dataset to explain both cross-section and time variability of prices considering also contextual variables like: room and hotel quality, offered services, restriction placed on prices or the spatial density of competitors or destination occupancy. Price is taken as response variable in a random-effect regression model and advance booking is an explicative variable. However, [Guizzardi et al., 2017] found that price trajectories can be seen as a stationary AR(1) process, suggesting that dynamics is present not only between successive arrival dates, but also along the advance booking, the role of which cannot then be properly assessed simply considering the booking lag as a covariate.

In this article, we overcome this issue by considering a panel-VAR model. The VAR setting enables us to consider the price trajectory as a multivariate endogenous variable, modelling also the interdependencies between different advance bookings, other than the serial correlation in each time series. The panel generalization of VAR models lets us build consistent GMM estimators considering the cross-sectional nature of the dataset. Moreover, the computation of the impulse response functions (IRF) enables us to interpret the effect of exogenous shocks (in our case holidays and fairs) in terms of forecasted price response. Thanks to this approach, we assess: if and how hotel managers account for demand patterns and exogenous shocks in their price competition on the on-line market; if and when they change from strategy to tattic along the advance booking.

The rest of the paper is organized as follows: in Section 2 we describe the dataset and the modellistic framework, introducing the general formulation of panel-VAR models, and discussing the techniques adopted for model selection and estimation. In Section 3 we present the main results, which include significant model coefficients and the relevant impulse response functions; in Section 4 we discuss briefly our findings, linking them to the existing literature about dynamic pricing, and present our concluding remarks and an outline of future extensions of this work.

2 Data and Methods

We consider a dataset consisting of best available rates (BARs) recorded every day at 00:00 AM for a panel of 107 hotels in Milan, from January, 1st to September, 30th, 2016. The data source was `Expedia.com`. For each arrival date, the room price has been recorded from 28 to 0 days of advance booking.

Let $i = 1,2,\ldots,N = 107$ index the hotel, $t = 1,2,\ldots,T = 274$ the arrival date and $k = 0,1,\ldots,K = 28$ the number of days of advance booking: let us denote the natural logarithm of the BAR for hotel i, arrival date t and advance booking k as $y_{it}^{(k)}$. We call *price trajectory* for the arrival date t at hotel i the vector of values $y_{it} = \{y_{it}^{(k)}\}_{k=0}^{28}$. In order to reduce dimensionality, we consider a sub-sample of the price trajectory, including a limited number of values of the advance booking to sample short, medium and long term with respect to the arrival date. In particular, we only consider lags = 0, 1, 7, 14, 21, 28, so that we reduce the dimension of the dependent variable from $K = 29$ to $K = 6$.

Given the panel nature of our dataset, and that for each hotel y_{it} is a vector time series, it is natural to consider the extension to panel-VAR models, which permits to take simultaneously into account the longitudinal and multivariate time series nature of the data. From our preliminary analysis, we establish that none of the hotel-specific VARs contains cointegration, so that no differencing is needed to achieve stationarity.

The panel-VAR model of order p reads ([Abrigo et al., 2016])

$$y_{it} = \sum_{j=1}^{p} y_{it-j}A_j + x_{it}B + u_i + \varepsilon_{it} \tag{1}$$

where y_{it} is a $(1\times k)$ dependent vector, x_{it} is a $(1\times l)$ vector of exogenous variables, u_i is a $(1\times K)$ vector of hotel-specific fixed effects and ε_i is a $(1\times K)$ vector of idiosyncratic errors, such that $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}'\varepsilon_{it}) = \Sigma$, $E(\varepsilon_{it}'\varepsilon_{is}) = 0$ for all $t > s$. A_j and B are, respectively, $(K\times K)$ and $(l\times K)$ matrices of parameters to be estimated.

The covariates x_{it} consist of four dummies indicating weekend nights (from Friday to Sunday), national holidays, fairs, and the month of August, during which business activity in Italy is considerably reduced. From a preliminary analysis of their marginal effects, we found that seasonal variability is not very evident, except for the months of August and September, during which prices are, respectively, lower and higher than the annual average. However, we can only link to tourism seasonality the lower price levels during August, while the higher fares in September are expectedly due to particular events, such as an high number of fairs during the month. Moreover, weekend nights appear to be associated to prices lower that the weekly average, including Fridays. Regarding short events displacing demand from its equilibrium state, fairs are associated to significantly higher room rates, while holidays do not seem to marginally affect room prices in Milan.

Estimation

We use the package `pvar` in `Stata`, described in [Abrigo et al., 2016], to specify the panel-VAR model in Eq. 1. Model estimation is conducted in a GMM framework, while model selection is based on model and moment selection criteria (MMSC).

Concerning the estimation of the coefficient matrices A and B in Eq. 1, the package follows a GMM approach which generalizes the bivariate procedure introduced by [Holtz-Eakin et al., 1988]. As suggested by [Arellano and Bover, 1995], we remove fixed effects through forward orthogonal deviation (FOD), which consists of subtracting the mean of only the future observations at any time and for any unit. This way, the variables transformed to remove fixed effects and the lagged dependent variables remain orthogonal, allowing to use the latter as instruments for estimation. The FOD procedure results in the following transformed variables, which are subsequently used for GMM estimation: $y^*_{it} = \frac{y_{it} - \overline{y_{it}}^+}{\sqrt{T_{it}/(T_{it}+1)}}$, where $\overline{y_{it}}^+$ is the mean of the future observations at time t, T_{it} is the number of remaining future observations at time t, and y^*_{it} is a $(1 \times K)$ vector $y^*_{it} = [y^{*1}_{it}\, y^{*2}_{it} \cdots y^{*K}_{it}]$.

Considering the model as a system, rather than equation-by-equation, makes it possible to achieve efficiency. In practice, this is obtained by writing the model equation in reduced form, $y^*_{it} = \tilde{y}^*_{it} A + \varepsilon^*_{it}$. The resulting GMM estimator for A is obtained stacking observations over panel and then over time, $\hat{A} = (\tilde{y}'^* z \hat{W} z' \tilde{y}^*)^{-1} (\tilde{y}'^* z \hat{W} z' \tilde{y}^*)$ where z are the available instruments, and $\hat{W}$ is a $(L \times L)$ weighing matrix, with $L = Kp + l$, assumed to be nonsingular, symmetric, positive semidefinite, and can be chosen to maximize efficiency according to [Hansen, 1982]. The GMM estimator is consistent if the orthogonality condition between instruments and idiosyncratic errors $E(z'\varepsilon) = 0$ holds and $rank\, E(\tilde{y}'^*_{it} z) = Kp + l$

Model selection

Panel-VAR model selection is based on the choice of a maximum lag for the set of instrumental variables and for the autoregressive model. For our dataset, we choose $P = 8$ as maximum autoregressive order; this choice is made considering that we may observe weekly dependencies which could be better caught by an AR(7) model, so that we want $p = 7$ to be among the eligible values. The selection of the maximum lag of the endogenous variables to be used as instruments is less arbitrary. With a view to testing the validity of the restrictions imposed by the postulated model, [Hansen, 1982] proposed an extension of the specification test introduced by [Sargan, 1959]. Such test requires over-identification, i.e. the number of available instruments must be larger than $kP+l$: since the exogenous variables provide the l valid instruments, we need to choose a number $q > p$ of lagged endogenous K-dimensional variables. Moreover, to avoid consistency loss due to endogeneity, instrumental variables are valid if incorrelated to the estimation variables, so that the smallest lag available as a valid instrument for y_{it-1} is y_{it-2}, and the smallest lag available for y_{it-2} is y_{it-3}. In practice, the minimum set of instruments for a panel-VAR(p) model is $[y_{it-2} \cdots y_{it-p-2}]$, which also implies that, while conducting the model selection, for every set of q instruments, we can only achieve over-identification up to $p = q-1$.

The order selection is based on model and moment selection criteria (MMSC) introduced by [Andrews and Lu, 2001] to include the number of parameters p, the number of instruments q and the test statistics $J(p,q,K)$ for over-identification defined by [Hansen, 1982].

3 Results

The procedure described in the previous section leads to the final selection of a panel VAR(2) with $[y_{it-2}, \cdots, y_{it-6}]$ as instruments to guarantee overidentification. The modulus of all the eigenvalues of the companion matrix lies inside the unit circle, so that the model is stable. Sargan's test does not reject the null hypothesis of correctly specified model, with a p-value = 0.080. Estimation results are shown in Table 3, only for the exogenous variables and significant coefficients of endogenous variables.

The only endogenous variables that are significant in an explanatory sense are prices at advance booking 0, 1 and 21. Prices at advance booking 0 and 1 are significant "predictors" of prices at larger advance bookings for arrival dates in the immediate future (which have been fixed several days or weeks earlier). This indicates that hotel managers not only practice revenue management in a dynamic pricing framework, but also that along the considered advance booking they make an implicit prediction of the last-minute occupation rate and of competitors' prices.

Concerning the exogenous variables, we can notice that periodic effects, such as month and day of week, have significant effects at all the considered lags of advance booking. Also fairs significantly affect price levels at all advance bookings along the trajectory, always with positive coefficients. On the other hand, while holidays do not show significant marginal impact on prices, a small positive holiday effect can be recognized for advance bookings in the close proximity to the arrival date. The weakness of this holiday effect and its limited extension along the trajectory, opposed to the stronger and always significant impact of fairs reflect the vocation of Milan as a business destination. We can also notice that the negative effect due to weekends is stronger for last-minute deals, while discounts associated to Summer holidays are stronger for early purchasers. Finally, the positive effect of fairs seems to be higher for larger advance booking. This highlights, for the examined period, a tendency of managers to be too optimistic about the increase in occupation rates stimulated by fairs, and then on feasible prices. This reflects a recent decline in the fair market in Milan, as noted by [Guizzardi, 2016].

Thanks to the possibility to compute impulse response functions (IRF), we can interpret the role of these effects also in a forecasting setting. Dynamic multipliers, i.e. the IRFs for exogenous variables, are shown in Fig. 1 and 2: while the IRF of the holiday dummy is flat on zero at all future lags, fairs clearly result in higher prices in the immediate future (i.e. the duration of the event), with a return to equilibrium around three days later. The estimated IRFs are shown with 95% Monte Carlo (500 replicates) confidence intervals.

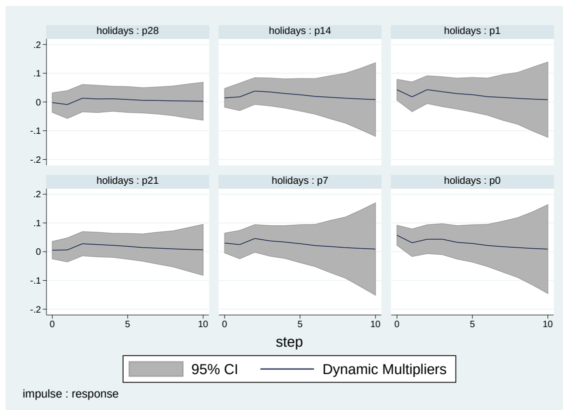

Fig. 1 Impulse response function on a 10 days horizon for holiday shocks. The gray shaded area marks the 95% confidence intervals obtained from 500 Monte Carlo replications.

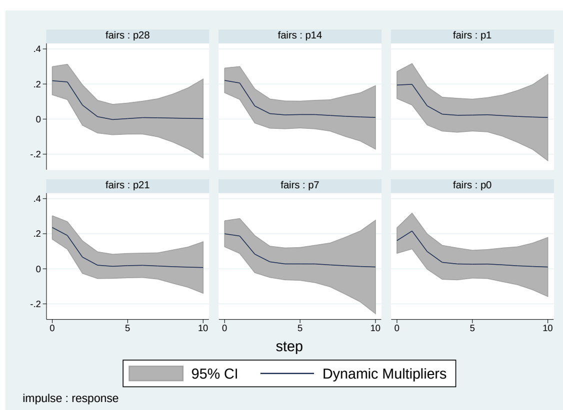

Fig. 2 Impulse response function on a 10 days horizon for fair shocks. The gray shaded area marks the 95% confidence intervals obtained from 500 Monte Carlo replications.

Table 1 Panel-VAR(2) significant coefficients.

y^0	Coef.	Std. Err.	p-value	y^1	Coef.	Std. Err.	p-value	y^7	Coef.	Std. Err.	p-value
y^0 lag=1	-	-	-	y^0 lag=1	-	-	-	y^0 lag=1	1.450	0.553	0.009
lag=2	0.310	0.114	0.007	lag=2	0.336	0.119	0.005	lag=2	0.234	0.107	0.029
y^1 lag=1	-1.514	0.702	0.031	y^1 lag=1	-1.830	0.710	0.01	y^1 lag=1	-2.126	0.673	0.002
lag=2	-	-	-	lag=2	-	-	-	lag=2	-	-	-
y^{21} lag=1	-	-	-	y^{21} lag=1	-	-	-	y^{21} lag=1	1.456	0.616	0.018
lag=2	-0.268	0.356	0.451	lag=2	-0.606	0.253	0.017	lag=2	-0.666	0.243	0.006
weekend	-0.199	0.015	0.000	weekend	-0.196	0.016	0.000	weekend	-0.144	0.016	0.000
august	-0.097	0.040	0.017	august	-0.121	0.042	0.004	august	-0.117	0.040	0.003
holidays	0.057	0.018	0.001	holidays	0.043	0.019	0.024	holidays	-	-	-
fairs	0.161	0.038	0.000	fairs	0.194	0.038	0.000	fairs	0.200	0.036	0.000
y^{14}	Coef.	Std. Err.	p-value	y^{21}	Coef.	Std. Err.	p-value	y^{28}	Coef.	Std. Err.	p-value
y^0 lag=1	1.056	0.519	0.041	y^0 lag=1	0.960	0.476	0.044	y^0 lag=1	1.172	0.572	0.04
lag=2	0.226	0.102	0.026	lag=2	0.208	0.092	0.024	lag=2	-	-	-
y^1 lag=1	-1.660	0.634	0.009	y^1 lag=1	-1.49	0.58	0.01	y^1 lag=1	-1.592	0.671	0.018
lag=2	-	-	-	lag=2	-	-	-	lag=2	-	-	-
y^{21} lag=1	1.880	0.531	0.001	y^{21} lag=1	1.950	0.645	0.044	y^{21} lag=1	2.078	0.591	0.000
lag=2	-0.763	0.239	0.001	lag=2	-0.634	0.361	0.024	lag=2	-0.854	0.233	0.000
weekends	-0.125	0.142	0.000	weekend	-0.123	0.013	0.000	weekend	-0.111	0.015	0.000
august	-0.130	0.038	0.001	august	0.141	0.035	0.000	august	-0.181	0.041	0.000
holidays	-	-	-	holidays	-	-	-	holidays	-	-	-
fairs	0.221	0.034	0.000	fairs	0.236	0.032	0.000	fairs	0.219	0.039	0.000

4 Discussion

In this work, we have shown that hotels of the high rating segment in Milan determine on-line BARs on the base of both their intentions about planning last-minute prices and the need to guarantee a certain stability of the price trajectory along the advance booking. In particular, the causality structure resulting from the VAR estimation suggests the significance of both last-minute and mid-term (three weeks in advance) room rates to explain the price dynamics. Prices in the mid-range of the trajectory (7-14 days) are not significant in an explanatory sense. The balance between tactic and strategic pricing holds up until short advance booking, where the tactic gains importance. Rates at advance booking 0 and 1, become negatively correlated with pricing strategies pursued at 21 days of advance booking. This negative correlation reflects tactic price adjustments in response to under/overbooking levels, caused by improper pricing choices in the mid-term, or an incorrect assessment of competitors on-line last-minute pricing. This behaviour is amplified by the prevalence of business travellers, who tend to make reservation with short advance booking while having low elasticity towards price.

We included four exogenous variables, all controlling for periods or events that modify the destinations business activity, so that we expect them to change the level of aggregate demand for the destination, occupation rates and, consequently, the pressure on firms to discount or increase prices. We find that fairs have a positive effect on prices, slightly higher for larger advance bookings, while the negative coefficients corresponding to weekends and the month of August call for a discount strategy. It is worth to notice that the negative effect due to Summer holidays is stronger for early purchasers, while discounts associated to weekends are stronger for last-minute deals. This suggest that managers have high confidence in planning pricing during August, while they give less weight to strategic pricing regarding weekends, given the unpredictability of weekend occupation rates, and lacking early information about: weather, events for leisure tourists and/or the concomitance with other significant events (also in the adjacent days) that interest the business segment. Also the effect of national holidays is significant only in short-term, while positive, for the reasons just mentioned regarding weekend pricing and for the fact that holidays are irregularly placed during the year. Overall, it appears that offer is not saturated until the advance booking is very small, so that pricing strategies defined three weeks in advance are only adjusted very close to the target date.

Concerning shock propagation, the impulse response functions for fairs and holidays show a different impact on price dynamics. While the response caused by a holiday tends to be zero at all future lags, fairs clearly result in higher prices in the immediate future (i.e. the average duration of the event), with a return to equilibrium around three days later.

In synthesis, we show that the actual strategies put into effect by the hotel managers reflect some of the basic principles of the online revenue management. Further development of this work will assess the question if, how, and to what extent, while practicing this mixture of tactical and strategic dynamic pricing, firms base their on-line pricing behaviour on the observation of prices published online by their competitors.

References

[Abrate and Viglia, 2016] Abrate, G. and Viglia, G. (2016). Strategic and tactical price decisions in hotel revenue management. *Tourism Management*, 55:123–132.

[Abrigo et al., 2016] Abrigo, M. R., Love, I., et al. (2016). Estimation of panel vector autoregression in stata. *Stata Journal*, 16(3):778–804.

[Andrews and Lu, 2001] Andrews, D. W. and Lu, B. (2001). Consistent model and moment selection procedures for gmm estimation with application to dynamic panel data models. *Journal of Econometrics*, 101(1):123–164.

[Arellano and Bover, 1995] Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of econometrics*, 68(1):29–51.

[Guizzardi, 2016] Guizzardi, A. (2016). Viaggi, fatturato e soddisfazione dei clienti del turismo daffari italiano nel 2015. *Turismo dAffari*.

[Guizzardi et al., 2017] Guizzardi, A., Pons, F. M. E., and Ranieri, E. (2017). Advance booking and hotel price variability online: Any opportunity for business customers? *International Journal of Hospitality Management*, 64:85–93.

[Guo et al., 2013] Guo, X., Ling, L., Yang, C., Li, Z., and Liang, L. (2013). Optimal pricing strategy based on market segmentation for service products using online reservation systems: An application to hotel rooms. *International Journal of Hospitality Management*, 35:274–281.

[Hansen, 1982] Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054.

[Holtz-Eakin et al., 1988] Holtz-Eakin, D., Newey, W., and Rosen, H. S. (1988). Estimating vector autoregressions with panel data. *Econometrica: Journal of the Econometric Society*, pages 1371–1395.

[Sargan, 1959] Sargan, J. D. (1959). The estimation of relationships with autocorrelated residuals by the use of instrumental variables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 91–105.

[Wang and Brennan, 2014] Wang, X. L. and Brennan, R. (2014). A framework for key account management and revenue management integration. *Industrial Marketing Management*, 43(7):1172–1181.