

The Number of Occupied Hotel Rooms A Time Series Model that Accounts for Constrained Capacity and Prices*

Kurt Brännäs and Jonas Nordström
Department of Economics, Umeå University
SE-90187 Umeå, Sweden

Abstract

The daily number of occupied hotel rooms in three large Swedish cities is modelled by an integer-valued and binomial autoregression. The model includes the capacity constraint and price variables are incorporated through the parameters of the model. The model implies a duration of hotel visit and an occupancy probability. We find that a 10 percent increase in the price level shortens the median duration of a hotel visit by approximately 6 percent during weekends and 8 percent during weekdays.

Key Words: Binomial, autoregression, estimation, demand analysis, rationed, price effect.

JEL Classification: C22, C25, C51, D21, L83.

Umeå Economic Studies 559, 2001

*The financial support from the Wallander Foundation is gratefully acknowledged. Xavier de Luna and seminar participants at the Stockholm School of Economics are thanked for comments on an earlier version of the paper.

1. Introduction

This paper models the daily number of occupied hotel rooms in three large Swedish cities by an integer-valued time series model. The model incorporates a total capacity constraint on the number of rooms as well as price and calendar time variables. The integer-valued nature of this type of data has previously been utilized by Brännäs, Hellström and Nordström (2001) and Brännäs and Nordström (2000), whose studies for international guests were based on monthly data at the national level, and could not utilize information on capacity constraints.

The present modelling exercise departs from the autoregressive binomial model of McKenzie (1985) in which the capacity constraint is an integral part. Our extension to this basic model makes it possible to account for time-variation in parameters and in the capacity constraint. It is a major benefit of the model that a survival function of hotel visits and an occupancy probability are implied, and that these can vary with, e.g., the price.¹ The model can be estimated from aggregate time series data using conventional least squares techniques.

The supply side for the hotel industry is characterized by the provision of hotel rooms, and as for most service industries the product is non-storable. Therefore, short term demand fluctuations are usually met by price changes, or by short term changes in the supplied capacity. Since prices usually are set some time before actual demand is known, i.e. under demand uncertainty, there may be periods when demand exceeds capacity.

There is a small literature on pricing (Scott, Satter and Highfill, 1995) and capacity setting (Carey, 1992) for hotels, while in the related area of the provision of hospital beds a larger number of studies has been undertaken (e.g., Friedman and Pauly, 1981, Mulligan, 1985). Deaton (1981) presents a theoretical model for rationed demand, and applies it to the British housing market. In addition to prices and expenditure Deaton's model includes the stock of the rationed good as a component in the demand system. Other studies that have considered price and capacity competition are among others, Davidson and Deneckere (1986) and Deneckere and Peck (1995). In a separate paper, we intend to develop and empirically test an adapted and economically stringent model to revisit the problem as considered here. Here, we consider estimation of single equations.

In Section 2 we introduce the basic time series model and demonstrate with motivations which changes are required to meet the characteristics of the hotel room market. Section 3 presents the data, while Section 4 introduces a nonlinear least squares estimator for

¹The alternative specification of Al-Osh and Alzaid (1991) is less well suited for our purposes. A conditional binomial specification would not imply the duration of hotel visit and the occupancy probability.

parameter estimation. The empirical findings are reported in Section 5. The final section summarizes our findings both with respect to the model and the number of occupied hotel rooms.

2. The Model

Occupied hotel room series are always subject to capacity constraints. When we wish to enforce this feature into a stochastic model, the binomial distribution comes naturally to mind. In a binomial model for the number of occupied rooms $y \in [0, N]$, N can be interpreted as a given capacity constraint. The other parameter that is required for the binomial distribution is the probability, θ , for a hotel room to be occupied. We may represent the number of occupied hotel rooms as

$$y = \sum_{i=1}^N u_i = \theta \circ N,$$

where the final expression is shorthand for the binomial thinning operation and is defined by the sum over independent and identically distributed 0 (representing an empty hotel room) and 1 (representing an occupied hotel room) random variables $u_i, i = 1, \dots, N$. The probability $\Pr(u_i = 1) = \theta$ so that $E(y) = \theta N$.

For the time series context, McKenzie (1985) suggests a first order autoregressive model with a binomial marginal distribution for y_t having parameters θ and N . The model is

$$y_t = \alpha \circ y_{t-1} + \beta \circ (N - y_{t-1}), \quad t = 2, \dots, T, \quad (1)$$

where the parameter $\beta = (1 - \alpha)\theta/(1 - \theta) \in (0, 1)$ and $\alpha \in (0, 1)$. Hence, the probability of a room being occupied is $\theta = \beta/[1 - (\alpha - \beta)]$, and depends both on the α and the β parameter.

Among the properties of the model in (1) we mention the conditional mean $E(y_t|y_{t-1}) = \alpha y_{t-1} + \beta(N - y_{t-1}) = (\alpha - \beta)y_{t-1} + \beta N$ and conditional variance $V(y_t|y_{t-1}) = \alpha(1 - \alpha)y_{t-1} + \beta(1 - \beta)(N - y_{t-1})$. Both measures are linear in past guest nights y_{t-1} as well as in the capacity constraint N . The model has an autoregressive conditional heteroskedasticity property (cf. ARCH). As $\text{Corr}(y_t, y_{t-1}) = (\alpha - \theta)/(1 - \theta)$ (McKenzie, 1985) this binomial model unlike other integer-valued autoregressive models of order one may have either negative or positive first order autocorrelation.²

² The binomial autoregressive model of Al-Osh and Alzaid (1991) can be written $y_t = S(y_{t-1}) + \varepsilon_t$. Here, $S(y)$ is the number of rooms remaining occupied out of y rooms, and ε_t is the number of check-ins. For given y , $S(y)$ follows a hypergeometric distribution implying that our α would vary with y . The interpretation and moment structure are different from those of our paper. For instance, this alternative model only has positive autocorrelation.

Note that under a constant θ assumption across n hotels, with varying capacity constraints $N_i, i = 1, \dots, n$, the cross-sectionally aggregated number of occupied hotel rooms remains binomially distributed and can be modelled by (1). The parameters are the same, but now $N = \sum_{i=1}^n N_i$. A nonconstant θ implies a nonbinomial distribution in the aggregate, and then the model representation (1) is not necessarily valid.

The expression for the conditional expectation suggests that we may interpret $\alpha - \beta$ as the probability of a hotel room remaining occupied an additional night, while β reflects the ability of hotels of filling free room capacity $(N - y_{t-1})$ with new guests. The hotel industry therefore likes to see a large $\alpha - \beta$ probability. A large β probability would minimize spare capacity but result in additional costs due to large guest turnover (small $\alpha - \beta$). Since, $1/[1 - (\alpha - \beta)]$ is the expected length of stay for a hotel guest, the expected income from a hotel guest is $p/[1 - (\alpha - \beta)]$, where p is the price for a single night.

To adapt the model for our empirical purposes we need to make a few changes to (1). We will, however, maintain the general model structure of (1). Our changes will enable us to utilize the conditional expectation expression directly for estimation, but will destroy the binomial property of the model. First, we wish to capture the varying capacity over days and an innocent change is then to introduce time variation in the total capacity variable N , i.e. we will use N_t . In addition, we make the assumption that N_t is pre-determined, in the sense that its value is set in advance of day t . Second, we wish to let, e.g., the price level have an effect on both α and β and therefore the parameters will vary with economic and calendar time variables. Note that as a consequence θ will also be time dependent. Third, one may expect dependence, e.g., between check-out and check-in decisions. Brännäs and Hellström (2001) demonstrated that weakening such independence assumptions will be of no consequence for the conditional mean, that we require for estimation. The effects only emerge in variance expressions and in higher order moments. Fourth, we will not enforce the interpretation of β in terms of the underlying α and θ parameters during estimation. We prefer obtaining estimates of θ from estimates of α and β . Therefore, we directly employ the restriction $\beta \in [0, 1]$ and write the model

$$y_t = \alpha_t \circ y_{t-1} + \beta_t \circ (N_t - y_{t-1}). \quad (2)$$

The explanatory variables in vectors \mathbf{x}_t and \mathbf{z}_t are introduced through logistic distribution functions (cf. the logit model)

$$\begin{aligned} \alpha_t &= 1/[1 + \exp(-\mathbf{x}_t\boldsymbol{\gamma})] \\ \beta_t &= 1/[1 + \exp(-\mathbf{z}_t\boldsymbol{\delta})], \end{aligned}$$

where $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are vectors of unknown parameters. Since $\alpha_t - \beta_t$ represents the probability of a hotel room remaining occupied another night, it appears reasonable to set $\mathbf{x}_t = \mathbf{z}_t$.

Using the definition of β (see above) we get the probability for a room being occupied at day t as

$$\theta_t = \beta_t / [1 - (\alpha_t - \beta_t)].$$

The parameters explain events occurring over the time interval $(t - 1, t]$ and depending on the context they may through the explanatory variables \mathbf{x} and \mathbf{z} depend on the state at time $t - 1$ or at t .

3. Data

The daily time series for the number of occupied hotel rooms (y_t) covers a seven day week for the three largest cities of Sweden; (by size) Stockholm, Gothenburg and Malmö, for the period January 1, 1993 to August 31, 1999 (source: Statistics Sweden; $T = 2434$ observations for each series). For the same period we also have access to the total capacity (N_t). Figure 2 gives statistics for capacity utilization in Stockholm and very clearly indicates a low utilization during Sundays. The plotted hotel price series in Figure 1 has a constant pattern within the month as it is constructed from monthly values for weekdays and weekends (Friday and Saturday), respectively.

From Figure 1, it seems evident that the hotel managers utilize intertemporal price discrimination to fill up the hotels in the short run. As the figure reveals there is a clear cyclical pattern within the week with higher prices from Sunday to Thursday and lower prices on Friday and Saturday when more leisure tourists visit the hotels. The price series also shows an annual cyclical pattern, with lower prices during July, which is the most common vacation month in Sweden. Empirically, we split the price variable into one for the weekdays and one for the weekends (Friday and Saturday).

In the estimated model we use four dummy variables. The Sunday number of guest nights is throughout small while the higher weekday price level is used. The Sunday dummy variable therefore controls for this irregularity. We also use a dummy variable for Monday and Thursday as these days appear to have slightly lower frequencies than the Tuesday and Wednesday nights. Another dummy variable catches the month of July, while the fourth accounts for the seven days before the new year and seven days after.

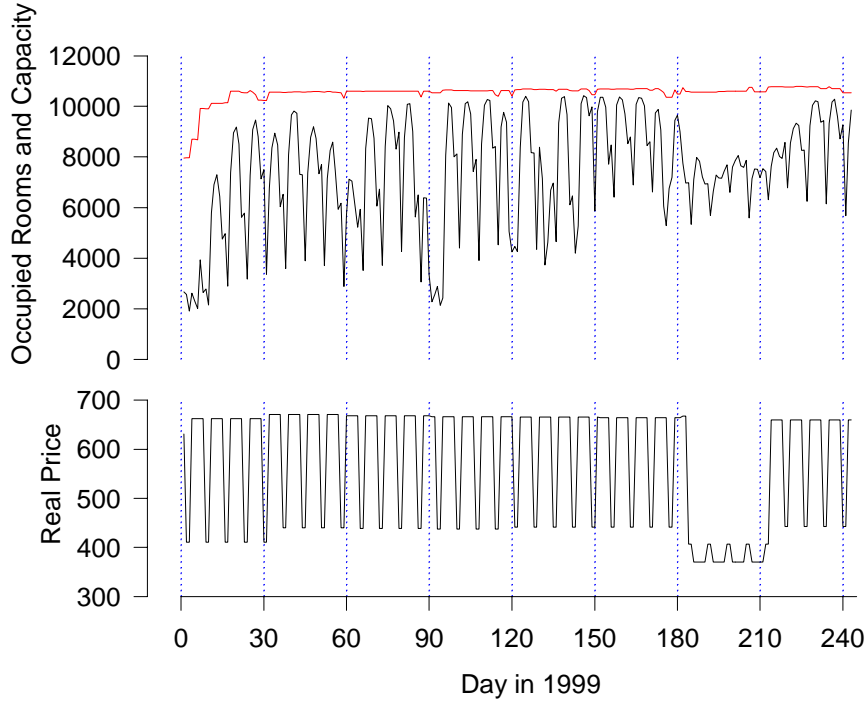


Figure 1: Number of occupied hotel rooms and total capacity in Stockholm (top) and the real hotel price (1986 prices, bottom) for 1999.

4. Estimation

To estimate the unknown parameters $\psi' = (\gamma', \delta')'$ we use a nonlinear conditional least squares (NCLS) estimator based on the prediction error

$$e_t = y_t - \alpha_{t-1}y_{t-1} - \beta_{t-1}(N_t - y_{t-1}).$$

The criterion to minimize is $S = \sum_{t=2}^T e_t^2$.

The covariance matrix of the estimator should account for the conditional heteroskedasticity. As we no longer maintain the binomial distribution assumption, and use an estimator based on the conditional first moment, some of the independence assumptions related to the thinning operations in (1) and (2) can be relaxed (cf. Brännäs and Hellström, 2001). One consequence is that an exact form for the conditional variance cannot be given. Therefore a covariance matrix estimator of the Eicker-White or Newey-West type is recommended.

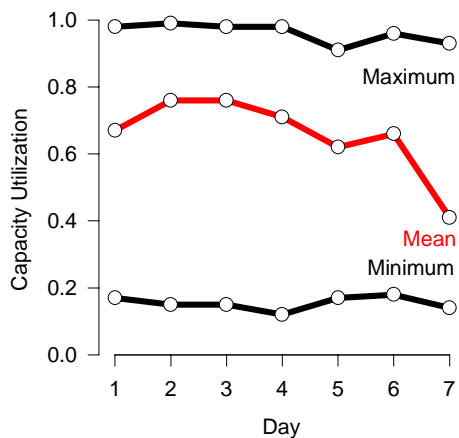


Figure 2: Capacity utilization for hotels in Stockholm along with minimum and maximum utilization vs. day (Day 1 corresponds to Monday; $T = 2434$).

If there is remaining serial correlation in the residuals an instrumental variable version of the nonlinear least squares estimator is called for. We obtain both estimators and use lagged explanatory variables as instruments in the instrumental variable estimator. The practical estimations reported in this paper were obtained by RATS 5.0, with the Newey-West covariance matrix estimator evaluated at 15 lags.

5. Results

Table 1 presents the estimation results. For all models there is significant serial correlation as judged by the Ljung-Box test statistic ($p = 0$). The largest autocorrelations are found at multiples of lag 7. Since the nonlinear least squares and the nonlinear instrumental variable estimators gave almost identical results, we only report the least squares estimates.

We first consider the price effect in the α -part. The expected negative effect is present for the price variable (Friday - Saturday) when there are mostly leisure tourists visiting the hotels. For the other days of the week the price effect is positive but insignificant for Stockholm and Gothenburg. Note, however, that the presence of the large and negative Sunday dummy variable can be expected to give a negative effect for that day. The same holds true for Monday and Thursday. The July dummy variable has a positive and

Table 1: Estimation results for models based on series in levels (s.e. in parenthesis).

	Stock- holm	Göthen- burg	Malmö	Stock- holm	Göthen- burg	Malmö
	α_t -part			β_t -part		
Constant	5.551 (0.394)	3.870 (0.397)	3.639 (0.457)	-2.646 (0.521)	-1.958 (0.424)	-2.053 (0.382)
Price \times 100	-0.687 (0.068)	-0.528 (0.071)	-0.549 (0.081)	0.117 (0.111)	-0.038 (0.099)	0.010 (0.097)
Friday-Saturday	0.204 (0.127)	0.062 (0.074)	0.340 (0.074)	0.240 (0.096)	0.133 (0.076)	0.096 (0.071)
Sunday-Thursday	-6.020 (0.508)	-4.127 (0.395)	-5.155 (0.549)	-0.873 (0.440)	-0.506 (0.267)	-0.195 (0.170)
Sunday	-4.456 (0.667)	-2.661 (0.412)	-4.322 (0.595)	1.150 (0.094)	1.034 (0.088)	1.266 (0.071)
Monday and Thursday	1.556 (0.484)	0.815 (0.349)	2.336 (0.374)	0.085 (0.514)	0.293 (0.423)	-0.677 (0.391)
July	-4.953 (0.505)	-3.299 (0.494)	-3.142 (0.642)	0.946 (0.490)	0.054 (0.383)	-0.046 (0.389)
New Year						
Average $\hat{\alpha}$	0.84 (0.22)	0.76 (0.24)	0.77 (0.26)			
Average $\hat{\beta}$	0.31 (0.35)	0.19 (0.06)	0.16 (0.05)			
Average $\hat{\theta}$	0.64 (0.30)	0.53 (0.27)	0.57 (0.30)			
R^2	0.81	0.79	0.76			
Ljung-Box, 15 lags	745	569	530			

Note: The price variables are not effective in July nor around the new year.

significant effect for all three cities, so that guests are expected to stay longer than during other periods of the year. For the period around the new year on the other hand guests stay a significantly shorter time.

For the β -part only the weekday price effect for Stockholm is significant and it is then of an unexpected positive sign when we interpret it as the ability of the hotels to fill free capacity. The Sunday effect is not throughout significant, while the Monday and Thursday effect is significant and positive. There is a mixed picture for the July and New Year dummies. Note that the impact of a variable in β is of very different magnitude depending on the size of $N_t - y_{t-1}$. For almost full capacity utilization there will hardly be any effect, but a larger one emerges for low utilization.

Since the probability of staying another night is given by $\alpha - \beta$ we illustrate the effect of a 10 percent price increase on the survival function,

$$\bar{F}(d|t-1) = \begin{cases} 1, & d = 0 \\ \prod_{i=t}^{t+d} (\alpha_i - \beta_i), & d = 1, 2, \dots \end{cases} ,$$

for check-ins during a Friday, Sunday, and Tuesday in the final full two weeks of the sample (August 1999) in Figure 3. For the three days the median durations are between 1.3 and 2.7 days. Using a market survey Holmberg and Westberg (2000) finds that inbound tourists stay on average 2.5 days in hotels, whereas domestic business visitors stay on average 2.0 days. The median durations in the figure are thus in rough agreement with the survey results.

The results in Figure 3 demonstrate that a price increase has the expected effect of shortening visit durations in Stockholm's hotels. The price increase shortens the median duration by approximately 0.15 days (6 percent) for Fridays, by 0.18 days (8 percent) for Tuesday, while there is practically no effect on the median for Sundays when the survival probability is lowest from the outset.

Figure 4 exhibits the model predictions for the daily occupancy probabilities during the final full week of August 1999. A very visible difference is the substantially higher utilization in Stockholm during Friday and Saturday. For this particular week, $\hat{\theta}$ and observed utilization are close for Monday-Thursday. Changing the price as in Figure 3 reveals that the occupancy probability is most affected during Friday-Saturday when the fraction of leisure travel peaks.

Another way of illustrating the price effect on y_t is to calculate the relative change in the conditional expectation. The effect $100(E_1 - E_0)/E_0$ is based on existing prices (p_t) for the conditional expectation E_0 and on a new price $1.01p_t$ for E_1 . Figure 5 gives the complete collection of relative effects for the different days and all three cities. The price

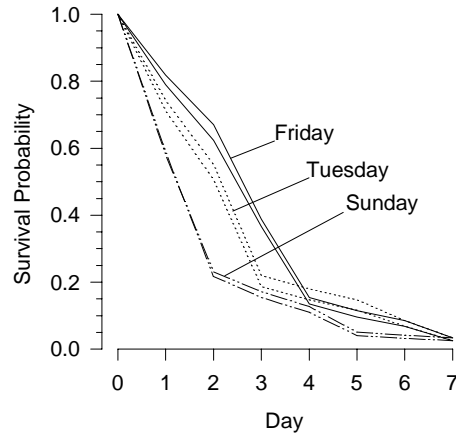


Figure 3: Survival functions for visit durations in hotels in Stockholm with check-ins during a Friday, a Sunday, and a Tuesday by the end of August 1999. The top lines are obtained using observed variables, while the lower lines correspond to a hotel price increase of 10 percent.

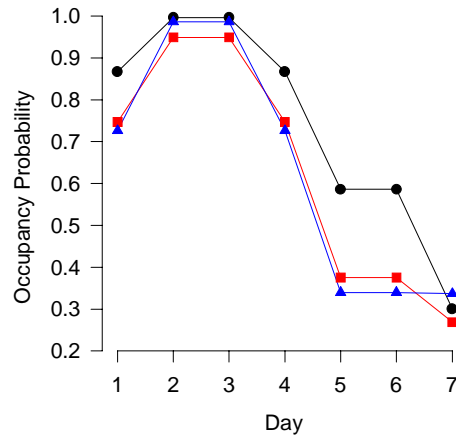


Figure 4: Predicted occupancy probabilities ($\hat{\theta}$) for the final full week of August 1999 (Day 1 corresponds to Monday; Stockholm, circle marker, Gothenburg, square, Malmö, triangle).

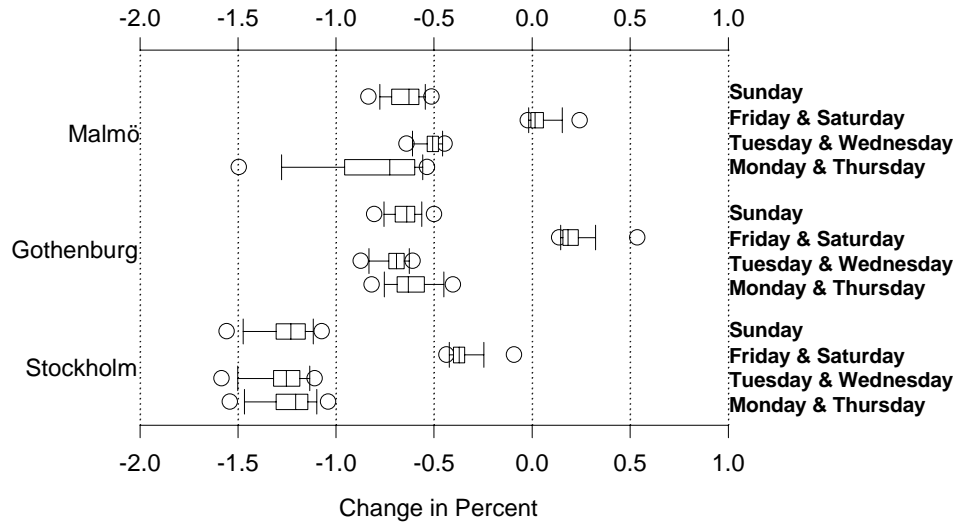


Figure 5: Box plots of changes (in percent) in the conditional mean of a one percent change in the real hotel price (based on all observations) for different days and cities.

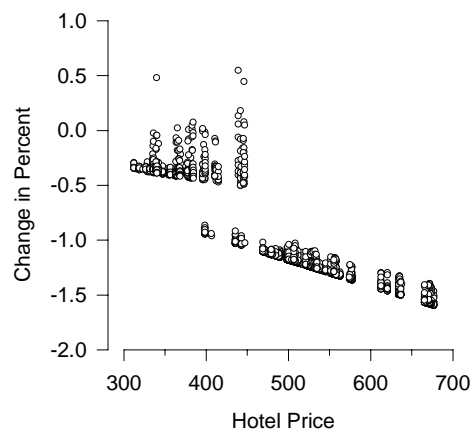


Figure 6: Change in percent against the hotel price for Stockholm and all observations.

effects are negative as expected for all but two instances. For Malmö the Friday-Saturday effect has a median just above zero, while for Gothenburg at the same days the effect is larger and significantly different from zero. Except for the Friday-Saturday effect in Malmö and Gothenburg the price effects are well in line with the results in Assarsson (1997), who estimates the short run price elasticity for hotel visits in Sweden to -1.3 , based on expenditure data from the national accounts. One reason to the positive price effects for Friday-Saturday in Malmö and Gothenburg is that the size of $N_t - y_{t-1}$ is large during these days. A reason for the lower price sensitivity for Friday and Saturday can be found in Figure 6, where the relative change in the conditional expectation (i.e. the own price elasticity) is plotted against the hotel price. As the figure reveals, the demand sensitivity depends on the level of the hotel price, and just as expected from the standard expression for the point elasticity we find a smaller percentage change in demand when the price is low.

6. Conclusion

We demonstrated how a basic integer-valued time series model could be adapted to include variable capacity and price variables. The model framework makes it possible to obtain survival functions for visit durations from aggregate time series data. For the considered time series the naturally incorporated capacity constraint turned out to be an important feature of the model. This was for instance found in the estimation of the occupancy probability, where the predicted and actual values were between 0.95 and 1.0 for Tuesday and Wednesday. The price effects we find on the duration and the conditional mean are well in line with results in other studies, although one could have expected to find larger effects for weekends compared to weekdays. This can, however, to some extent be explained by the lower price level during weekends.

References

- Al-Osh, M.A. and Alzaid, A.A. (1991). Binomial Autoregressive Moving Average Models, *Communications in Statistics: Stochastic Models*, 7, 261-282.
- Assarsson, B. (1997). Efterfrågan på tjänster i Sverige: Beräkning av efterfrågeelasticiteter och sysselsättningseffekter, SOU 1997:17, Appendix 2, Norstedts, Stockholm.
- Brännäs, K. and Hellström, J. (2001). Generalized Integer-Valued Autoregression, To appear in *Econometric Reviews*.
- Brännäs, K., Hellström, J. and Nordström, J. (2001). A New Approach to Modelling and Forecasting Monthly Guest Nights in Hotels, To appear in *International Journal of Forecasting*.
- Brännäs, K. and Nordström, J. (2000). A Bivariate Integer-Valued Allocation Model for Guest Nights in Swedish Hotels and Cottages, Umeå Economic Studies 540.
- Carey, K. (1992). Optimal Hotel Capacity: The Case of Barbados, *Social and Economic Studies*, 41, 103-126.
- Davidson, C. and Deneckere, R. (1986). Long-run Competition in Capacity, Short-run Competition in Price, and the Cournot Model, *RAND Journal of Economics*, 17, 404-415.
- Deaton, A.S. (1981). Theoretical and Empirical Approaches to Consumer Demand under Rationing, In *Essays in The Theory and Measurement of Consumer Behavior in Honour of Sir Richard Stone* (Deaton, A.S., ed.), Cambridge University Press, Cambridge.
- Deneckere, R. and Peck, J. (1995). Competition over Price and Service Rate when Demand is Stochastic: A Strategic Analysis, *RAND Journal of Economics*, 26, 148-162.
- Friedman, B. and Pauly, M. (1981). Cost Functions for the Service Firm with Variable Quality and Stochastic Demand: The Case of Hospitals, *The Review of Economics and Statistics*, LXIII, 620-624.
- Holmberg, I. and Westberg, M. (2000). Turistdatabasen – TDB: En statistisk utvärdering, Department of Statistics, School of Economics and Commercial Law, Göteborg University.
- McKenzie, E. (1985). Some Simple Models for Discrete Variate Time Series, *Water Resources Bulletin*, 21, 645-650.

- Mulligan, J. (1985). The Stochastic Determinants of hospital Bed Supply, *Journal of Health Economics*, 4, 177-185.
- Scott, R., Satter, E. and Highfill, J. (1995). A Hotel Capacity Utilization Model, *Journal of Economics*, XXI, 101-105.