Diversion ratios from share data and transition matrix estimates

Svetlana Golovanova[♦]

Eduardo Pontual Ribeiro[♦♦]

Abstract

We propose the use of ecological inference forecasting methods to estimate a transition matrix from (unconditional) market share data. This can be linked to calculating diversion ratios for firms in a market. While requiring high frequency market share data, it provides an alternative to calculate diversion ratios without price information and without imposing IIA assumptions as in a logit model. We use the method in two ways. First, to calculate diversion ratios in a market. The latter can be used in merger analysis. Second, to calculate the effect of exogenous changes in a market as a forecast tool alternatively to least squares multivariate prediction (VAR) models.

Keywords: transition matrix; diversion ratio; counterfactual

PRELIMINARY AND INCOMPLETE

This version July 30, 2024

[♦] National Research University – Higher School of Economics, Nizhny Novgorod, Russia.
[♦♦] Institute of Economics - Federal University of Rio de Janeiro (IE/UFRJ), Brazil.

Introduction

A key information for merger analysis is the substitution pattern between potential competitors in a merger (Davies and Garcés, 2010). This informs on the closeness of competition and can be used to delimit relevant markets and, through diversion ratios, inform on possible price effects of mergers, though upward pricing pressure statistics (UPP) or a full fledged simulation (Valleti and Zenger, 2021, e.g.).

Diversion ratios may require significant amount of data as they require estimating a full-fledged demand system. Alternatively, churning data (information on customer switching) may be used to understand consumer switching from firm A to firm B[1]. Under strict independence of irrelevant alternatives assumption in a Logit demand model diversion ratios could be calculated using only share data(with the requirement of specifying the share of an outside good, or external diversion (Moresi and Zenger, 2018)

We explore an alternative to calculating diversion ratios when only (unconditional) market share data is available. The method is based on recent developments in ecological forecasting methods that estimate transition matrix parameters imposing the coefficient restrictions in a linear programming problem Pavia and Romero (2022), Romero (2020).

Estimation methods are not new, at least in the political science literature, but have not seen use in the economic literature and no use in the industrial organization literature.

Two recent papers closer to ours are Qiu et al. (2024) and Conlon and Mortimer (2024). Qiu et al. (2024) explore the issues when only churning data (win/lose) data is available to calculate the transition probabilities and infer diversion rates. Their contribution include a theoretical result that explain the differences and similarities between diversion ratios based on transition probabilities, the common 'loss diversion ratio' and the unconditional shares. Conlon and Mortimer (2024) propose advanced estimation methods for a transition matrix when some information on transitions are available (but not all transitions), basically from merging parties only. Our method explores the situation when there is no transition/churning/win-loss data. And the results can be incorporated in the Qiu et al. (2024) equations, without imposing the independence of irrelevant alternatives assumption.

The method is used to calculate diversion ratios based on website visit data for the market of online travel agencies (OTAs) in Russia. The choice of Russia is due to the second application, namely, the estimation of the effect of the changes in market structure from March 2022, when the largest and international platforms stopped providing services on Russian lodging and international payment methods were blocked in the country. We explore the use of forecasting methods to estimate effects of interventions using autoregressive methods, extending the ideas of Bianchi and Salvati (2015) for a multivariate setting. Our presentation is descriptive and we believe informative, while the interpretation of causal effects require additional assumptions compared to the those required for estimation and forecasting.

[1] E.g, telecom merger case CASE M.7758-HUTCHISON 3G ITALY / WIND / JV, DG COMP, 2016.

1. Model and methods

Assume there is a market with two products (1 and 2). To use transition matrices, we must consider an outside good, to provide consumers an alternative when deciding not to consume products 1 or 2 or consumers that start to consume the goods. This is similar to the Logit demand model (e.g. Davies and Garcés, 2010), where an outside good must be defined so that the consumption alternatives are all considered. Accordingly, define i=0,1,2 to be the ‘goods’ in the system.

Let s_i be the (output) share of good *I*, $s_i=q_i/(q_0+q_1+q_2)$. Consider S_t the vector of shares at time *t* $S_t=[s_{0t}, s_{1t}, s_{2t}]^T$. We understand this as an unconditional distribution of market shares. Of course $\Sigma_j s_{jt}=1$.

The transition matrix P is a logical tool that updates the market share unconditional distribution $S_t= \Pi S_{t-1}$ for all *t=1,…*

$$\begin{bmatrix} s_{1t} \\ s_{2t} \\ s_{0t} \end{bmatrix} = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{10} \\ \pi_{21} & \pi_{22} & \pi_{20} \\ \pi_{01} & \pi_{02} & \pi_{00} \end{bmatrix} \begin{bmatrix} s_{1t-1} \\ s_{2t-1} \\ s_{0t-1} \end{bmatrix} \tag{1}$$

with the restriction that $\Sigma_j\pi_{jk}=1$ (columns add to one). If we take the shares as probabilities for different choices/goods, π_{jk}=P(choice *j* at time *t* | choice *k* at time *t-1*).

Estimating the matrix *Π* requires assumptions. The most common is the one of homogeneity, namely, that the matrix *Π* is the same for all *t* in a sample, as in the ecological inference forecasting literature (Petropolous et al., 2022, section 2.10.4). Assuming the π_{ij} are constant over time, two period large panel samples or even a time series sample of S_t can be used for estimation. Pavia and Romero (2020) and Romero et al. (2020) specified a linear programming model to incorporate restrictions. Their strategy is to convert the problem from shares to quantities, assuming first, a constant over time total quantity.[2] The empirical model estimates

$$\begin{aligned} q_{1t} &= \pi_{11}q_{1t-1}+ \pi_{12}q_{2t-1}+ \pi_{10}q_{0t-1}+\varepsilon_{3t} \\ q_{2t} &= \pi_{21}q_{1t-1}+ \pi_{22}q_{2t-1}+ \pi_{20}q_{0t-1}+\varepsilon_{2t} \\ q_{0t} &= \pi_{01}q_{1t-1}+ \pi_{02}q_{2t-1}+ \pi_{00}q_{0t-1}+\varepsilon_{1t} \end{aligned} \tag{2}$$

minimizing $\Sigma_j \Sigma_t |\varepsilon_{jt}|$. The link between minimum absolute value estimation and linear programming has been disseminated in the econometric literature through the quantile regression model (Koenker and Bassett, 1978). The linear programming method used for estimation allows for easier inclusion of parameter bounds restrictions and logical restrictions from shares.

There are a number of restrictions in the estimated model, as we started to discuss above. First and second, $\Sigma_j s_{jt}=1$ observed dependent variables across equations add to one and shares are non-negative ($s_{jt}\geq 0$). Third, $\Sigma_j \pi_{jk}=1$, columns add to one. Fourth $0\leq \pi_{jk}\leq 1$ for all *j* and *k*. The methods are packaged in an R suite `lphom`.

[2] This assumption is relaxed in the empirical implementation.

Inference is based on resampling methods. Romero et al. (2020) proposes a generated random bootstrap method. This method is time consuming and have proven impractical to the data set used, with more than 300 observations and at least five alternatives. We estimate, alternatively, a leave-one out jackknife variance estimation method. While earlier literature showed that for least absolute deviation (LAD) the leave one out jackknife estimate would be inconsistent, Bianchi and Salvani (2015) argue that simulations show the inconsistency to be of second order. Note that the associated confidence intervals can be asymmetric, as the bootstrap/jackknife parameters estimates can bunch at the limits of the parameter space.[3]

A key issue is to set the outside good size, or the size of the total market. Such issue is common in models that estimate Logit demand models. We use as total market size the maximum number of transactions in the sample. This figure can be adjusted for different assumptions of the market size.

1.1. From transition matrix to diversion ratio

The estimated transition matrix coefficients can be used to calculate diversion ratios, that measure closeness of competition. The diversion ratio is defined as $D_{jk}=-\Delta q_k/\Delta q_j$. Other definitions highlight the role of prices in the quantity variation: $D_{jk}=- (\partial q_k/\partial p_j)/ (\partial q_j/\partial p_j)$ for demand functions $Q_j(p_1, \ldots, p_n, z)$, *e.g.*, Davies and Garcés (2010). Diversion ratio measures the proportion of output that diverts from good (brand) *j* to good *k*. In the case of single unit purchase consumers, the proportion of consumers that stop consuming good *j* (exit consumers) and replace it with good *k* (entering consumers).

Using data from customers switching from firm 1 to firm 2 (churning data), one can calculate transition probabilities as in the system (1) above (Miller and Sheu, 2020, e.g.). The most common definition, using transition matrix ideas, is $D_{12}=\pi_{21}/(1- \pi_{11})$, e.g.,

Under the assumption of a simple Logit demand model (Davies and Garcés, 2010) the diversion ratio would be calculated as $D_{12}=s_2/(1- s_1)$. One may interpret the Logit demand diversion ratio, based on market shares, as a long run, stable, transition matrix outcome. It is well known that in 'steady state' $S^*=\lim_{m\to\infty} \Pi^m$, where the S* matrix has equal columns, each equal to the so called long run market shares. The method proposed here does not impose the transition matrix and associated diversion to be this long run outcome.

Qiu et al. (2024) provide an explanation for transition matrix coefficients to differ from the observed market shares, namely, switching costs, that generate state dependence. If there were no adjustment costs, the conditional transition probabilities from a utility maximization problem would be similar to unconditional choice probabilities, namely, observed market shares. They also show that in a full system of conditional choice probabilities, as in (1), the Diversion ratio can be precisely calculated, generalizing the Chu and Schwarz (2016) example that churning may differ from actual diversion. We use the simpler, better know diversion from conditional probabilities (known as loss diversion

[3] For example, consider, for a full sample estimate of 3.5 the following jackknife parameter estimates 3.2, 4, 0, 0, 5. The 20% percentil is 0 and the 80% percentil is 5.

in their paper), while the results can be extended [future version] to the full expression in their equation (5).

1.2. The transition matrix as a forecasting tool.

The model can be used to calculate expected market shares. The transition matrix model (1) could be written as a vector auto-regression (VAR) for time series. There are restrictions to straight forward estimation of the VAR. First, a practical one: as shares add up to one, the explanatory variables in the VAR in each equation (lagged market shares) are perfectly collinear. Second, the limited nature of the dependent variable. Parameter restrictions, leave one equation out and variable elimination as in the estimation of linearized almost ideal demand models (Davies and Garces, 2010, e.g.) or Translog cost functions for factor demand models (Berndt, 1990) are required.

restrictions should be considered. Additional parameter First $\Sigma_j \pi_{ji}=1$ (columns add to one). And second $0 \leq \pi_{ji} \leq 1$ for all *i* and *j*. . Fourth, as the shares add up to one, across equations *j=0,…,J*, the (reduced form) errors in a VAR are restricted $\Sigma_j \varepsilon_{jt}=0$ and not simultaneously independent, so the errors are not simultaneously independent.

Taking into account that $\Sigma_j \pi_{ji}=1$ (columns parameters add to one) the VAR can be rearranged as

$$
\begin{aligned}
s_{1t} &= \beta_{11}s_{1t-1} + \beta_{12}s_{2t-1} + \beta_{10} + \varepsilon_{3t} \\
s_{2t} &= \beta_{21}s_{1t-1} + \beta_{22}s_{2t-1} + \beta_{20} + \varepsilon_{2t} \\
s_{0t} &= \beta_{01}s_{1t-1} + \beta_{02}s_{2t-1} + \beta_{00} + \varepsilon_{1t}
\end{aligned}
$$

where $\beta_{11}=(\pi_{11}-\pi_{10})$, $\beta_{12}=(\pi_{12}-\pi_{10})$, $\beta_{21}=(\pi_{21}-\pi_{20})$, $\beta_{22}=(\pi_{22}-\pi_{20})$, $\beta_{10}=\pi_{10}$, $\beta_{20}=\pi_{20}$ and so forth. Original parameters may recovered using $\pi_{11} = \beta_{11} + \beta_{10}$, $\pi_{12} = \beta_{12} + \beta_{10}$ and accordingly for other parameters. The constant term appears as $s_{0t-1}=1-s_{1t-1}- s_{2t-1}$

In the VAR model above, one of the equations can be omitted as the dependent variable across equations add to one, and the residuals add to zero across equations for a given date. This is well known from the LAIDS factor demand models systems. Also, given the parameter restrictions, the equations of the parameters are such that one of the equations does not have free parameters. (e.g.Berndt, 2011). The omitted equation parameters may be recovered from the other parameters. If we omit the third equation, $\pi_{0j} = 1 - (\pi_{1j} + \pi_{2j}) = 1 - [(\beta_{1j} + \beta_{2j}) + (\beta_{10} + \beta_{20})]$ for j=1,2. For convenience, the omitted equation may be the one for the outside good.

The estimated system can be rewritten as

$$
\begin{aligned}
s_{1t} &= \beta_{11}s_{1t-1} + \beta_{12}s_{2t-1} + \beta_{10} + \varepsilon_{3t} \\
s_{2t} &= \beta_{21}s_{1t-1} + \beta_{22}s_{2t-1} + \beta_{20} + \varepsilon_{2t}
\end{aligned}
$$

The parameters of the VAR are still restricted in their possible values, namely, $-1 \leq \beta_{ij} \leq 1$ for i,j=1,2 and $0 \leq \beta_{i0} \leq 1$. Such inequality/bounds restrictions can be imposed by altering the functional form of the equation to a non-linear model where transformations of unbounded parameters estimated by least squares would generate parameters that correspond to the restrictions[4]. For example if b_{10} is the estimated parameter from least squares, $c_{11}=\exp(b_{10})/[1+\exp(b_{10})]$, guarantees that the result is bounded between zero and one. Shortcuts such as could work in principle, but appear an adhoc way to impose the coefficient restrictions.[5] Instead of using such ad hoc adjustments, we use the ecological forecasting method of Romero et al (2020) as described in the previous section.

1.3. Forecasts as counterfactuals for impact analysis

The methods above can be used to provide forecasts. These forecasts can have, under restrictions, help identify causal effects of an exogenous intervention, that would lead to model parameter changes. Forecasts can be used to estimate counterfactuals, had the parameters not changed. The causal effect would be identified by the difference between the observed and the counterfactual estimated by the model. The idea of using a prediction model to calculate counterfactual is standard in event studies in finance (MacKinnon, 1999). If the causal effect is a level shift in the series, e.g., and average treatment effect, the parameters, of say, an autoregressive model will shift, as in Menchetti et al. (2023).

In a simple differences-in-differences model the control group dynamics acts as the counterfactual forecast, under the common trend assumption. In addition, if the dynamics are stable over time, i.e., before and after the intervention, the counterfactual could be calculated either from the control group or from the treatment group itself.

The dynamics of the data can be modelled as a VAR, the generalization of an autoregressive model. The intervention should generate parameter changes. The estimates take into account that the selected intervention is exogenous and have spillovers across studied treated units, thus requiring a system of equations as model (1). Note that the shares add up to one restriction imply that the forecast error between observed and counterfactual outcomes cancel each other across equations.

For our specific application, the methods are applied in a different fashion than most event studies. Given that the intervention occurred in the beginning of our sample, we estimate the model using a post intervention data set and predict the earlier part of the data, as in Figure 1. Under stationarity of the data and within estimation period stable parameters, the time frame for estimation should not effect (beyond sampling variation) the estimated coefficients. This idea has a very old history in econometrics dating back to Chows' predictive failure structural break test (e.g., Judge et al., 1988). As in the event study literature, the chosen (start) date of the estimation sample may be far from the actual intervention to allow adjustment of consumer behavior to the new dynamics.

Figure1 – Time estimation strategy

t=1	t=t* (intervention)	T=T
Prediction sample	Estimation sample	

[4] https://www.stata.com/support/faqs/statistics/linear-regression-with-interval-constraints/

[5] Actual application yield negative or greater than one coefficients in the transition matrix, as some of the parameters are calculated as differences between coefficient sums and one.

2. Applications.

The applications use similar data sets, namely, daily number of visits on online travel agencies (OTAs) in Russia, collected from Similarweb.com. The advantage of the data is that is high frequency, with an observation per day. The disadvantage is that the data does not record if the visit resulted in a sale. We use only reservation websites, excluding websites that mostly provide price comparisons.

Up to 2022, the market participants were either multinational firms such as Booking.com or AirBnb.com, or local firms, such as Yandex Travel, followed by much smaller alternatives namely Onetwotravel, Ostrovok and Sutochno. Yandex Travel is the travel agency provider a Yandex, an ecosystem with browser, search engine, social network and off line services such as taxi and delivery.

Table 1. Online Travel Agencies Platforms studies main characteristics

	Geography	Platform services, other than accommodation booking	Website language	2022 Jan. visits share	2023 Jan. visits share
Booking.com	Worldwide (198 countries)	air tickets, Car rental, taxi booking, tour booking	43 languages	55.0	12.5
Ostrovok.ru	Worldwide (220 countries)	No	24 languages	2.3	21.7
Travel.yandex.ru	Worldwide	air, railway and bus tickets; tour booking	16 languages	31.1	34.0
Onetwotrip.com	Worldwide	air, railway and bus tickets; tour booking; car rental	9 languages	6.2	5.7
Sutochno.ru	Focus on Russia and CIS countries (47 countries total)	No	Russian, English	2.8	10.3
Tvil.ru	Russia, Abkhazia, Belarus and Georgia	No	Russian		6.4
101hotels.com	Worldwide (64 countries)	No	Russian		7.6

Booking - Starting from March, 5 2022 Booking.com does not provide the service of booking hotels located on the territory of Russia and Belarus, while booking hotels in other countries through the platform is not restricted to Russians. Tvil and 101hotels – data collected from July 1st, 2022. Agoda, AirBnB, CBooking, Expedia, Hotels, Trip with less than 1% of market share.

In the beginning of 2022, many foreign companies including a number of digital platforms that provide accommodation booking services stopped their business in Russia. Booking.com, a world leading online travel agency for lodging reservations and other travel products, is one of them, halting reservations of Russian units and not accepting Russian payment methods, but still accessible in the country. Before the exit, on March 5th, 2022, the company was a dominant provider of the services in Russia. Airbnb.com and Agoda.com are two more platforms that stopped providing their services in Russia on the same date. However, their market shares were much smaller. The decision of the international platforms not to deal with hotels and private owners offering apartments for lease on the territory of Russia resulted in a sharp redistribution of the market between competing providers of the same services.

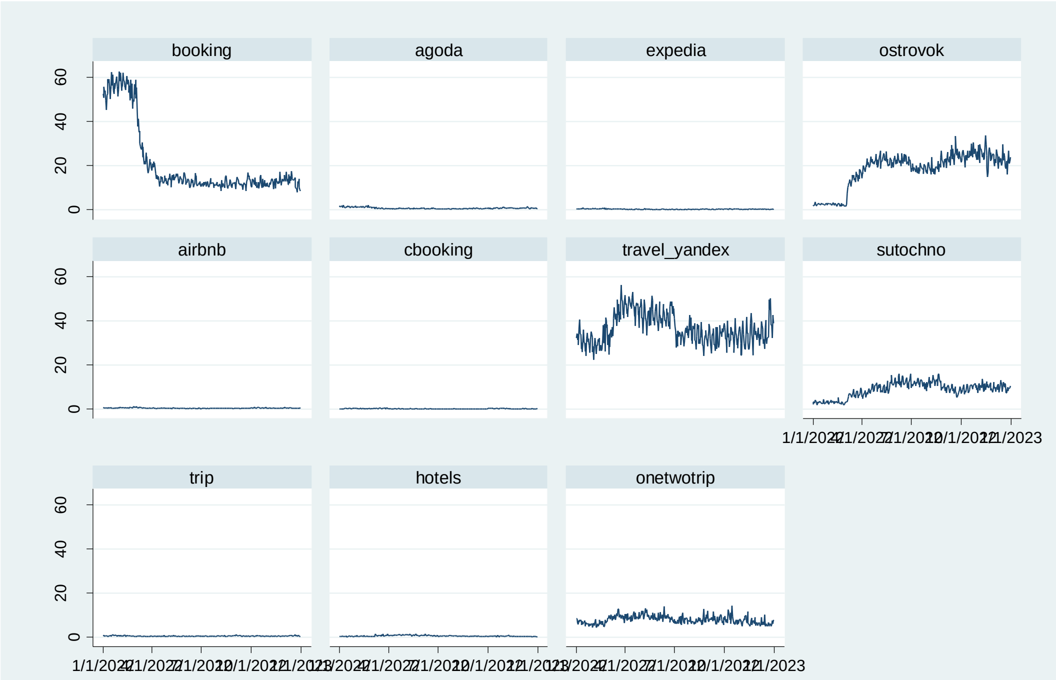
At the same date Visa and Mastercard suspended their activities in Russia. It was one more shock on the Russian accommodation booking market as this restricted the ability of Russian residents to do international payments for online services. The use of their credit cards outside Russia became not possible as well. This led to increase in domestic tourist trips and accommodation booking on the territory of Russia despite the absence of restrictions for lodging reservations abroad by hotel booking platforms.

The daily data on the number of visits of users with Russian IP addresses is collected for the largest platforms that provide accommodation booking services (online travel agencies) in Russia. These are the top platforms that were recommended as substitutes to Booking.com and Airbnb.com by different online Russian forums right after the exit of the companies. We consider only direct providers of the booking services that intermediate both reservations and payments. The websites that overview offers from different online travel agencies and provide links to their web sites for further booking (like tripadviser.com, trivago.ru, hotellok.ru etc.), i.e., price comparison websites, are not considered in the study as mentioned.

The data is collected for the period from January 1st, 2022 to December 31st, 2023. The key characteristics of 7 largest platforms in our database are presented in the Table 1. The rest 3 platforms are Agoda.com (left Russia in March 5th 2022), Trip.ru and Hotels.ru with total share less than 3% of the total market measured as the maximum number of visits in the sample.

Figure 1 provides an overview of the industry in 2022. We see that Booking and Travel Yandex basically divided the market up to Booking delisting of Russian accommodation units and payment restrictions. We see an increase in local alterantives such as Ostrovok and Sutochno.

Figure 1 – Market shares - Online Travel Agencies (OTAs) – Russia – 2022.



Note Shares for firms with information – excludes outside good.

2.1.Estimating diversion ratios from market share data and a transition matrix.

For the application we use only data for 2023, where the effect of the international platforms exiting the market has been already incorporated in the market dynamics. Only the websites with at least 1% market share on average are included. These are Booking, Ostrovok; Travel.Yandex; Sutochno; Onetwotrip; Tvil. The outside good is calculated based on the maximum number of visits in a day in an year (in the summer). The outside category includes smaller platforms.

The estimates transition matrix is presented on Table 2

Table 2 - Parameter Estimates – Transition Matrix – Six largest Online Travel Platforms – Russia – 2023 (as %)

	Previous Date Platfom visited						
Current date platform visited (below)	**Booking**	**Ostrovok**	**Yandex Travel**	**Sutochno**	**OneTwoTrip**	**Tvil**	**Other**
Booking	40.9	6.7	5.5	0.0	13.7	0.0	1.7
	[40 - 42.3]	[6.3 - 7.1]	[5 - 5.8]	[0 - 0]	[10.7 - 16.6]	[0 - 0]	[1.6 - 1.7]
Ostrovok	29.7	53.2	10.6	2.9	0.0	17.7	1.8
	[28.5 - 31.5]	[52.4 - 53.8]	[10.3 - 11.8]	[2.3 - 3.3]	[0 - 0.3]	[16.2 - 18.5]	[1.7 - 1.8]
Yandex Travel	17.6	23.6	41.5	39.2	41.6	0.0	4.6
	[16.3 - 18.6]	[23.2 - 24.6]	[40.6 - 42]	[38.5 - 40.1]	[38.6 - 44.5]	[0 - 0]	[4.5 - 4.7]
Sutochno	0.1	9.0	7.1	52.2	22.5	22.2	0.0
	[0 - 0.6]	[8.7 - 9.6]	[6.5 - 7.3]	[52.1 - 52.3]	[22.2 - 23.1]	[21.9 - 22.4]	[0 - 0]
Onetwotravel	11.4	3.8	4.7	3.0	22.2	0.0	0.0

	[10.6 - 11.5]	[3.7 - 4.3]	[4.7 - 4.9]	[2.7 - 3.3]	[21.9 - 22.8]	[0 - 0]	[0 - 0]
Tvil	0.0	3.8	4.2	2.7	0.0	60.1	0.0
	[0 - 0]	[2.8 - 4.7]	[3.5 - 4.8]	[2.4 - 3.3]	[0 - 0]	[59.2 - 61.5]	[0 - 0]
Other	0.3	0.0	26.3	0.0	0.0	0.0	92.0
	[0 - 1.4]	[0 - 0]	[26.1 - 26.6]	[0 - 0]	[0 - 0]	[0 - 0]	[92 - 92.1]

Note: estimates from model (2). 10% confidence intervals in brackets, based on leave one out jackknife estimates. Columns add to 100.

The interpretation of the coefficients of the transition matrix is standard. Starting from the first column, say, of 100 views of Booking in one period, 40.9 views were observed for Booking the next period, while 29.7 shifted to Ostrovok and 17,6 moved to Travel Yandex. The first column results suggest that Sutochno and Tvil, as well as the ‘other’ category are not close substitutes, as there does not seem to be shifts from Booking views to these platforms/alternatives.

There seems to be persistence in choices, as the diagonal coefficients are the column largest and at least 40%, except for OneTwoTravel. The estimates transition coefficients are not symmetric. For example, the chance of moving from Booking to Ostrovok is 29,7% while the chance of moving from Ostrovok to Booking (second column) is much smaller, at 6.7%. There is a number of zero coefficients mostly to/from ‘other’.

The confidence intervals appear tight around the coefficient estimates. They are not symmetric as they are based on sampling estimates and the coefficient restrictions (namely, non-negativity and column adding to one) appear to be binding in many cases.

Table 2 calculates diversion ratios, based on the well know formula of $\pi_{jk}/(1-\pi_{kk})$, i.e., relative to the chance of *not* continuing in choice *k* at time *t*-1, what is the chance of moving to choice *j* at time *t*. The first column suggests, that from 100 visits that used Booking in a period and do not visit it in the next period, 50% moved to Ostrovok, 30% to Travel Yandex, none moved to Sutochno and 19% moved to OneTwoTravel. According to these estimates, Ostrovok and Tavel Yandex are the closest competitors to Booking.

Also note the sum of off-diagonal items in each column. The smaller the sum (but for the last line “outside good”, or “not buying”) means that the firm is more isolated from competition as the firm loses less consumers to competition when such consumers decide not to purchase. This outside good appears not to influence the market as the diversion is zero, except for Travel Yandex.

Table 3 - Diversion Ratios – Six largest Online Travel Platforms – Russia, 2023

		From collumn Site					
		Booking	Ostrovok	TY	Sutochno	Onetwotravel	Tvil
To	Booking		0.14	0.09	0.00	0.18	0.00
	Ostrovok	0.50		0.18	0.06	0.00	0.44
	TravYandex	0.30	0.50		0.82	0.53	0.00

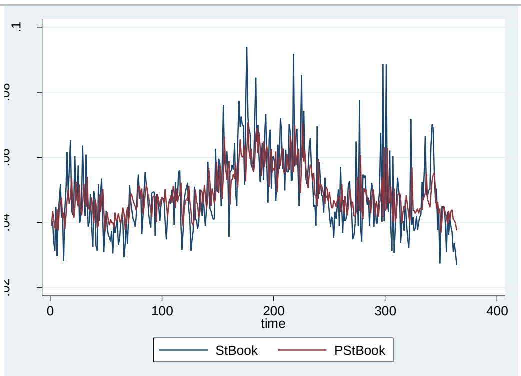
	Sutochno	0.00	0.19	0.12		0.29	0.56
	Onetwotrav	0.19	0.08	0.08	0.06		0.00
	Tvil	0.00	0.08	0.07	0.06	0.00	
	Aggregate Diversion	1.00	1.00	0.55	1.00	1.00	1.00

Note: based on Table 2. TY – Travel Yandex. Aggregate diversion is the sum of column diversion rates.

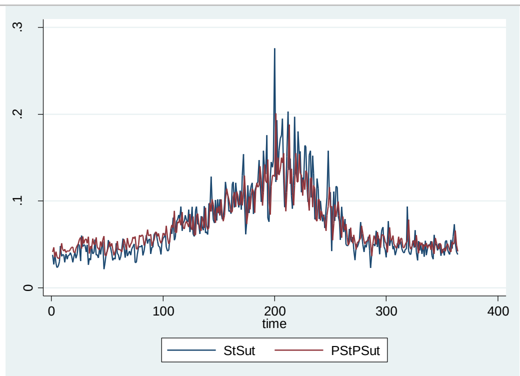
A more detailed analysis of the service may provide an explanation to this result. In contrast to other OTA platforms in the analysis, TY is a part of ecosystem Yandex (108 services), a Russian equivalent to Google ecosystem. The TY service was introduced by Yandex in the end of 2020 (relatively new) with the focus on cross-platform comparison of offers on air tickets form other aggregators (including Onetwotravel, for example). There are reasons to believe that hotel booking service is considered by users like a secondary one for this platform. As seen in table 1, the comeptitors Ostravok Sutochno, Onetwotravel and Tvil focus mainly on lodging.

Overall, the transition matrix estimated in Table 2 provides very good predictions of the market shares (including the outside good), as may be seen on Figure 2.

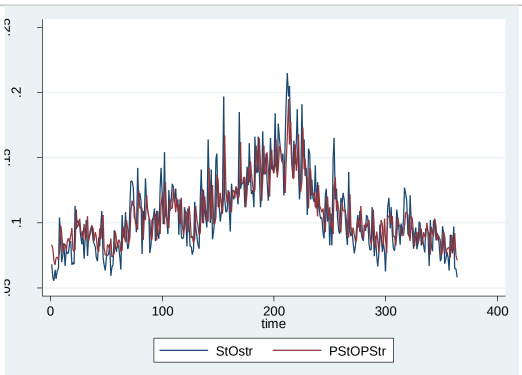
Figure 2. Actual and predicted market shares of OTA platforms in Russia, 2023



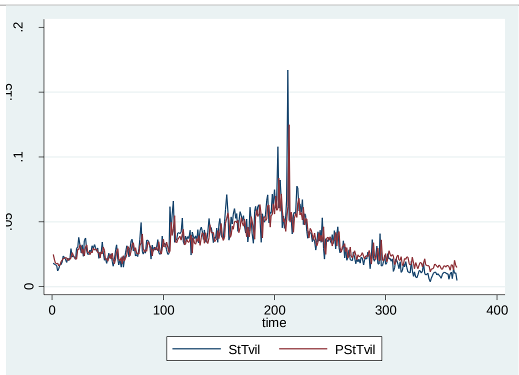
Booking



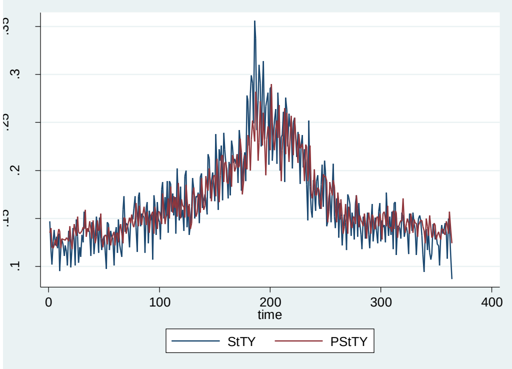
Sutochno



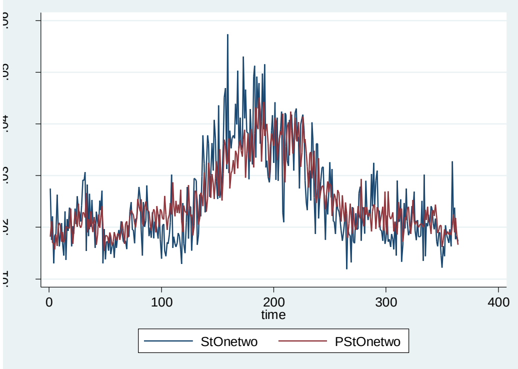
Ostrovok



Tvil



TravelYandex



Onetwotrip

Note: blue line (St__) are market shares, calculated including the outside good. Red (PSt__) are predicted shares.

The results from the transition matrix are quite different from the simple share-based diversion. The results are presented on table 4, for the case of shares – including the outside good – for January 2023. First, there are no alternatives with zero diversion, as in table 3. Second, the relative importance of alternatives do not change, as expected. This is in sharp contrast to table 3 above. While Ostrovok has the highest diversion from Booking (first column of table 3), but a very low diversion from Sutochno, this result is not available for the estimates the use the share-based diversion.

Table 4 – Share based Diversion Ratios – Six largest Online Travel Platforms – Russia, 2023

		From collumn Site					
		Booking	Ostrovok	TY	Sutochno	Onetwotravel	Tvil
To row Site	Booking		0.06	0.06	0.06	0.06	0.06
	Ostrovok	0.10		0.11	0.10	0.10	0.10
	TravYandex	0.15	0.16		0.15	0.15	0.15
	Sutochno	0.05	0.05	0.05		0.05	0.05
	Onetwotrav	0.03	0.03	0.03	0.03		0.02
	Tvil	0.03	0.03	0.03	0.03	0.03	
	Aggregate Diversion	0.35	0.33	0.28	0.36	0.37	0.37

Note: based on Table 1 market shares, expanded for t. TY – Travel Yandex. Aggregate diversion is the sum of column diversion rates. Market shares used: Booking: 5.4%, Ostrovok: 9.3%, Travel Yandex: 14.5%, Sutochno: 4.4%, Onetwotravel: 2.4%, Tvil: 2.8%, outside option: 61.2%

2.2. The transition matrix as a prediction models. The effect of market leader exit.

In early March 2022, Booking exited the Russian market. At the same time, Visa and Mastercard started not accepting transactions from Russian issued cards or Russian acquirers. This is a unique opportunity to understand closeness of competition as the market leader left the market for reasons exogenous to the company. [why are so many visits to Booking still after March 2022? International travels of Russian nationals?]

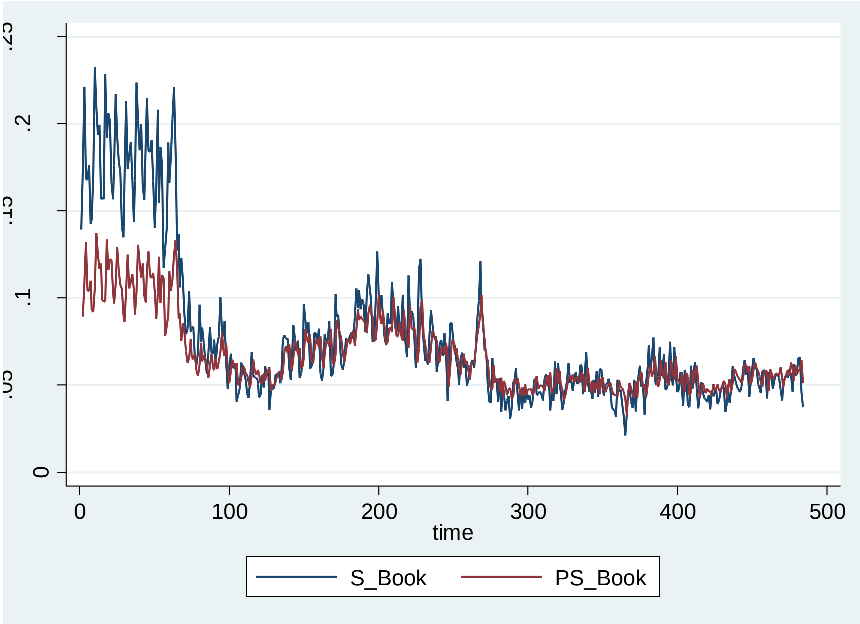
These are the calculated market shares using the sum of visits for the 11 platforms with data From January 1st, 2022 to April 30th, 2023. The total market is the maximum number of visits in the sample (which happens to be in the summer). It is clear from figure X that booking had the largest number of visits in the market, at the start of 2022, followed by Yandex Travel. Even at the start of 2022 some OTAs (including price comparison sites) were not popular, with less than one percent share, as AirBnB, Expedia and Hotels.com. After the exit of booking and other international websites, there is a large surge in visits of local alternatives, such as Ostrovok and Sutochno. Yandex Travel does see an increase in visits as well. The market is basically divided between by the end of 2022: Yandex Travel, Ostrovok, Sutochno and OneTwoTrip. We aggregate the others and the difference of the visits to the maximum of daily visits as 'others' or the outside good.

The parameter estimates are less of interest here, as we want to explore the fact that the exit of Booking was unexpected. This generates a natural experiment that we explore here for calculating counterfactuals. As explained earlier, we estimate the model for data from May 1st, 2022 to April 30th, 2023 (a full year of data) and use the estimates for one step ahead forecasts for the period January 1st-April 30th, 2022. We allow two months for the market to accommodate the exit from Booking in early March, i.e., our estimation sample begins in May.

Note that the underlying assumption of the counterfactuals is that the change in the market brought about the basic exit of Booking from reservations in Russia and payments methods restrictions led to a change in the dynamics of the visits to the websites. As the model is based on the dynamics of this platforms and (potential) competitors, aggregate effects but the intervention such as seasonality, income shifts should cancel out across firms, under homogeneity of income effects. At the same time, losses from booking should translate to gain in shares of some alternative, including the 'other'/outside good.

Figure 1 provide a visual representation of the results for Booking website. We see, again, as above, that the model provides close predictions within estimation sample, and show a clear gap in observed and counterfactual shares for the first part of the sample.

Figure 3 – Booking.com observed (blue) and predicted shares (red), based on model (1).



Note: Blue – observed shares; Red – predicted shares; Shares include the outside good. Estimating data from May 2022.

We aggregated the observed market share, the predicted market share and the forecast error by month. Within estimation sample we expect that, in thirty days of a month the predictions are unbiased, i.e., zero mean forecast error. This is indeed the case. Looking at the estimates for months from April on, we see that the model predict market shares well. On the other hand, the predicted and observed shares differ significantly. While the average share (including the outside good) is 18% in January and February, the predicted share (based on behavior of the market from April onwards) is only 11%.

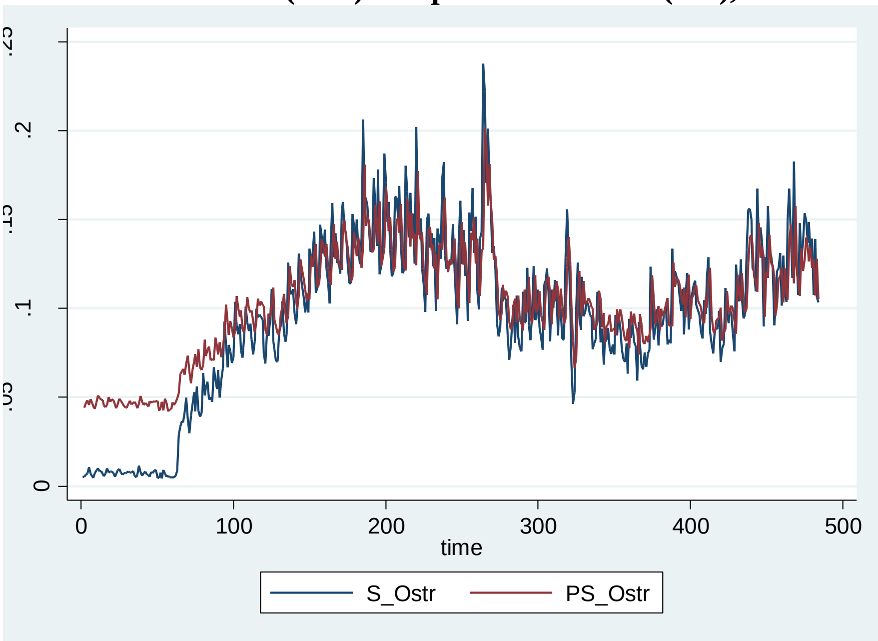
Table 5 – Observed and predicted shares – Booking.com

month	Observed Share	Predicted Share	Mean Error
1	0.18	0.11	0.07
2	0.18	0.11	0.07
3	0.11	0.08	0.03
4	0.06	0.06	0.00
5	0.06	0.06	0.00
6	0.07	0.07	0.00
7	0.09	0.09	0.01
8	0.08	0.08	0.00
9	0.07	0.07	0.00
10	0.05	0.05	-0.00
11	0.05	0.05	-0.00
12	0.05	0.05	-0.00

Note: intervention in early March. Estimating sample May, 2022-April 2023.

A mirror effect of the exit of the OTA with the highest market share is the Ostrovok website. The counterfactual market share for the company in January and February is much higher, were the consumer dynamics from April onwards in place earlier that year. The company experienced a 4p.p. increase, on average, in the market share that can be attributed to the Booking exit. Again, notice that the forecast error for each month after the structural change in the market is very small, at most a percentage point.

Figure 4 – Ostrovok observed (blue) and predicted shares (red), based on model (1).



Note: Blue – observed shares; Red – predicted shares; Shares include the outside good. Estimating data from May 2022.

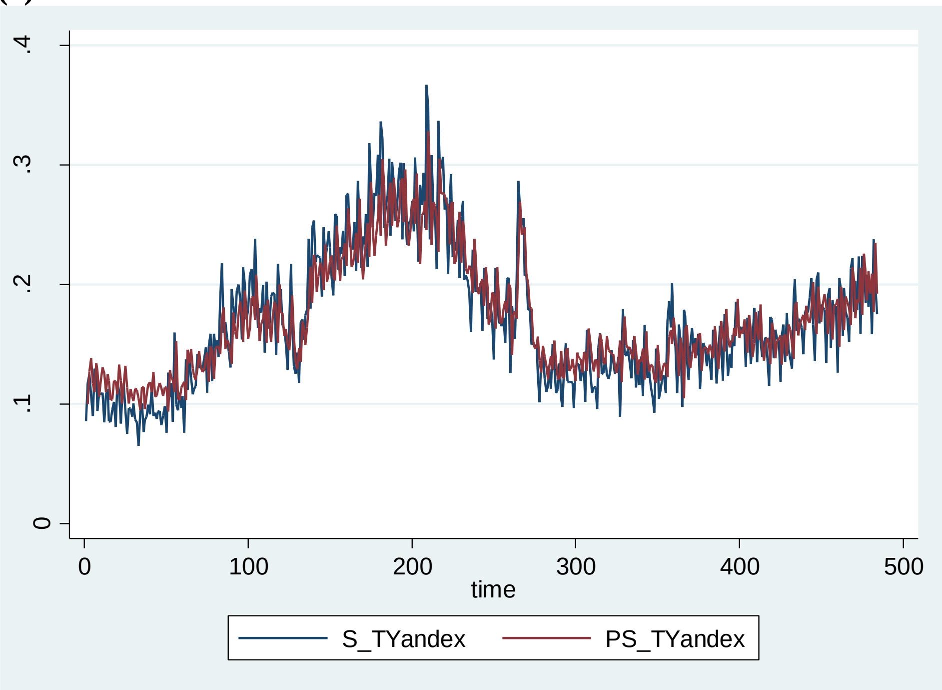
Table 6 – Observed and predicted shares – Ostrovok.com

month	Avg share	Avg predicted share	difference
1	0.01	0.05	-0.04
2	0.01	0.05	-0.04
3	0.04	0.07	-0.03
4	0.08	0.09	-0.01
5	0.10	0.10	-0.01
6	0.13	0.13	0.00
7	0.15	0.14	0.01
8	0.14	0.14	0.01
9	0.14	0.14	0.01
10	0.09	0.10	-0.01
11	0.10	0.10	-0.00
12	0.08	0.09	-0.01

Note:intervention: early March 2022. Estimating sample – May 2022-. Table based on 2022 data.

On the other hand, the model suggests that Yandex Travel was not much affected by the the structural change in the market. The prediction error is much smaller in this case, suggesting that the change in dynamics given the restrictions on Booking did not alter the behavior of the visits to the platform. One argument is that the website is more or less insulated from competition, given its emphasis on other travel related services, such as air and rail reservations, compared to other websites. Recall that, post intervention, using data for 2023, Yandex had the largest aggregate diversion rate, suggesting that when users did not access the website they chose not to visit any other OTA.

Figure 5 – Travel Yandex observed (blue) and predicted shares (red), based on model (1).

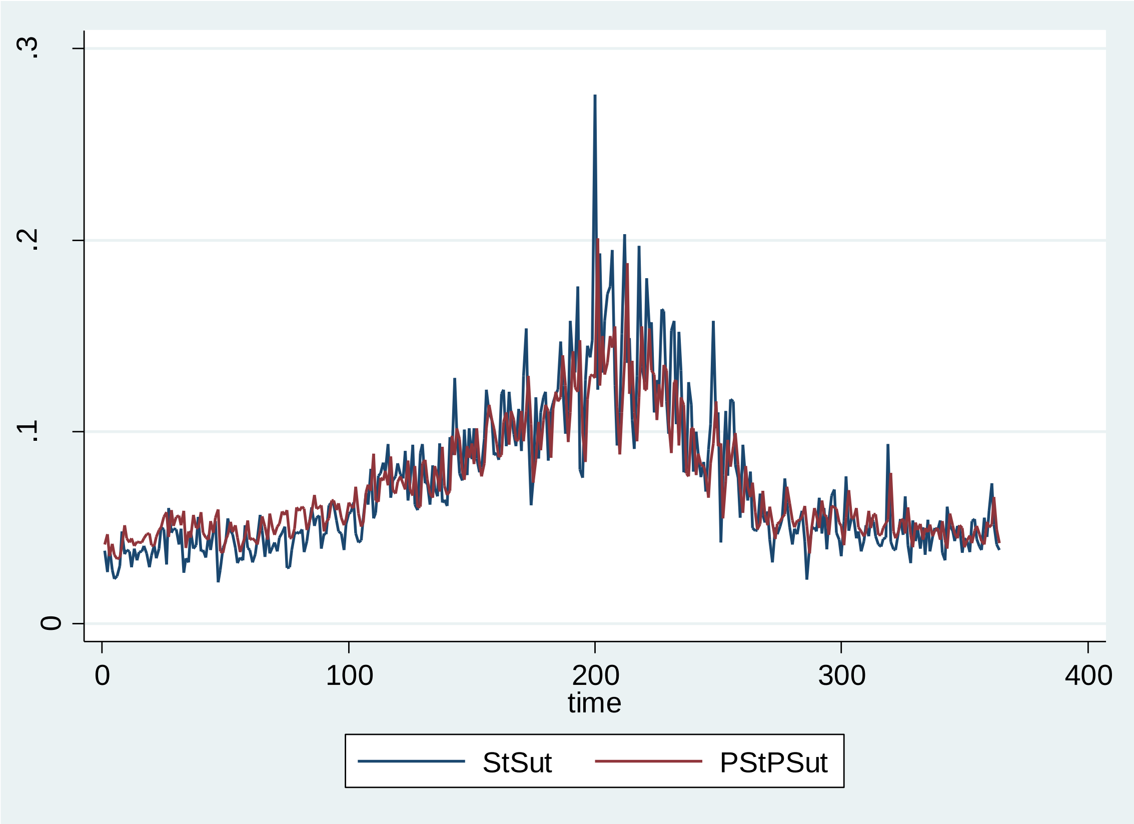


Note: Blue – observed shares; Red – predicted shares; Shares include the outside good.
Estimating data from May 2022.

Table 7 – Observed and predicted shares – Travel Yandex

month	Observed Share	Predicted Share	Mean Error
1	0.10	0.12	-0.01
2	0.10	0.11	-0.02
3	0.14	0.14	-0.00
4	0.19	0.17	0.02
5	0.19	0.18	0.01
6	0.25	0.23	0.01
7	0.28	0.27	0.01
8	0.24	0.24	0.00
9	0.19	0.19	-0.00
10	0.13	0.14	-0.01
11	0.13	0.14	-0.01
12	0.14	0.14	-0.00

Figure 6 – Sutochno observed (blue) and predicted shares (red), based on model (1).



Note: Blue – observed shares; Red – predicted shares; Shares include the outside good.
Estimating data from May 2022.

Table 8 – Observed and predicted shares – Sutochno

month	Observed Share	Predicted Share	Mean Error
1	0.01	0.02	-0.01
2	0.01	0.02	-0.01
3	0.02	0.03	-0.01
4	0.03	0.04	-0.01
5	0.04	0.05	-0.00
6	0.07	0.07	0.00
7	0.09	0.09	0.01
8	0.09	0.08	0.01
9	0.05	0.06	-0.01
10	0.04	0.04	-0.00
11	0.04	0.04	-0.00
12	0.04	0.04	-0.00

Concluding Comments [draft]

Propose the use of a new method to the problem of estimating diversion ratios from share data. Uses methods developed in the political science literature

Method avoids the IIA assumption and no adjustment costs/persistence in consumption, providing flexible, data driven diversion ratios.

Method requires no data beyond market shares. Method requires high frequency data (or many data points with stable transitions) and still requires assumptions about the outside good/total market size.

We use the method in two applications: One: estimating a transition matrix (with confidence intervals) and diversion ratio to measure closeness of competition. Second: a prediction model to infer the effect of the exit of a significant player to the shares of others (natural experiment on the changes in the closeness of competition).

We apply the method to online travel agencies (OTA) digital platforms visit daily data, for 2022 and 2023. We use 2023 data to explore the transition matrix and associated diversion. We use the 2022 data to explore the effect of the exit of internation websites, basically the market leader Booking.com from the market to reservations in Russia. Our results indicate than the second largest platform, the local YandexTravel benefitted little from the exit of Booking, apart from becoming the largest, in a proportional shift, while other local platforms absorbed a disproportionate share of the the diverted demand.

REFERENCES

Bianchi A., Salvati N. (2015) Asymptotic Properties and Variance Estimators of the M-quantile Regression Coefficients Estimators, *Communications in Statistics - Theory and Methods*, 44(11), 2416-2429 doi:10.1080/03610926.2013.791375

Chen, Y, Schwartz M. (2016) Churn Versus Diversion in Antitrust: An Illustrative Model, *Economica*,. 83(332), 564-583

Conlon, C., Mortimer, J.H. (2021), Empirical properties of diversion ratios. *The RAND Journal of Economics*, 52: 693-726. https://doi.org/10.1111/1756-2171.12388

Davies, O., Garcés, E. (2010) *Quantitative techniques for competition and antitrust analysis*, Princeton:PUP.

Huang D. and Rojas C. (2013) The Outside Good Bias in Logit Models of Demand with Aggregate Data. *Economics Bulletin*, 33(1), 198-206.

Koenker, R.,Bassett, G. (1978) Regression Quantiles, *Econometrica*, 46 (1), 33-50

Menchetti, F., Cipollini, F., and Mealli, F. (2023) Combining counterfactual outcomes and ARIMA models for policy evaluation, *The Econometrics Journal*, 26(1) 1–24, https://doi.org/10.1093/ectj/utac024

Miller, N.H., Sheu, G. (2021) Quantitative Methods for Evaluating the Unilateral Effects of Mergers. *Review of Industrial Organization* 58, 143–177

Moresi, S. Zenger, H. (2018) Recapture ratios in merger analysis, *Economics Letters*, 170 (1) 136-138,

Pavía, J. M., and Romero, R. (2022). Improving Estimates Accuracy of Voter Transitions. Two New Algorithms for Ecological Inference Based on Linear Programming. *Sociological Methods & Research*, 0(0). https://doi.org/10.1177/00491241221092725

Petropoulous, F. et al. (2022) Forecasting: theory and practice, *International Journal of Forecasting* 38(3) 705-871.

Qiu, Y.J., Sawada, M. and Sheu, G. (2024), Win/Loss Data and Consumer Switching Costs: Measuring Diversion Ratios and the Impact of Mergers. *Journal of Industrial Economics* 72(1) 327-355. https://doi.org/10.1111/joie.12371

Romero, R., et al. (2020) Assessing uncertainty of voter transitions estimated from aggregated data. Application to the 2017 French presidential election, *Journal of Applied Statistics*, 47(13-15), 2711-2736, https://doi.org/10.1080/02664763.2020.1804842

Valletti, T. and Zenger, H. (2021) Mergers with Differentiated Products: Where Do We Stand?, *Review of Industrial Organization* 58(1), 179-212.