Black Market Performance: Illegal Trade in Beijing License Plates*

Øystein Daljord† Mandy Hu ‡ Guillaume Pouliot § Junji Xiao ¶

August 14, 2019

Abstract

Black markets are often thought to create welfare gains by allocating restricted goods to agents who value them the most. The illegal nature of black markets however create transaction costs that impede trade. We estimate the incentives to trade in the black market for license plates that emerged following the recent rationing of new car sales in Beijing by lottery. Under weak assumptions on car preferences, we use optimal transport methods and comprehensive data on car sales to estimate that at least 12% of the quota is illegally traded . We develop a simple market equilibrium model which allows us to infer bounds on transaction costs, such as potential legal liabilities, and transaction prices, given the estimated market size. The inferred transaction prices are on par with those reported anecdotally in the daily news. We find that up to 66% of the gains from trade are lost to transaction costs. The size of the transaction costs points to potentially severe market frictions.

*A warm thanks to Shanjun Li for comments and data. Thanks to Carl Mela, Luxi Shen, Avner Strulov-Shlain, and Thomas Wollmann for comments. Thanks to Charles Ahlstrom for research assistance.

†Chicago Booth School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA. E-mail: Oeystein.Daljord@chicagobooth.edu. Web: faculty.chicagobooth.edu/oystein.daljord .

‡CUHK, Department of Marketing, Room 1105, 11/F, Cheng Yu Tung Building, 12 Chak Cheung Street Shatin, N.T., Hong Kong. E-mail: mandyhu@baf.cuhk.edu.hk. Web: bschool.cuhk.edu.hk/staff/hu-mandy-mantian/.

§Harris School of Public Policy, University of Chicago,1155 E 60th St, Chicago, IL 60637 USA E-mail: guillaumepouliot@uchicago.edu. Web: sites.google.com/site/guillaumeallairepouliot.

¶Department of Economics, UTS Business School, University of Technology Sydney 14-28 Ultimo Rd., Ultimo, NSW2007, Australia. E-mail: junji.xiao@uts.edu.au. Web: uts.edu.au/staff/junji.xiao.

Keywords: informal economy/underground economy and optimal transport partial identification semiparametric and nonparametric methods

JEL codes: E26, D450, P230, C140

1 Introduction

The informal economy, the trade in goods and services that goes undetected in official statistics, makes up an estimated one sixth of the GDP in the world economy (Schneider et al. (2010)). A subset of the informal economy are black markets: markets where goods and services are traded illegally. These markets are often considered undesirable *per se*, e.g., markets for hard drugs, but do not always have undesirable consequences. Black markets are thought to enhance welfare by reducing the distortionary impact of rationing and price regulations on the allocation of resources (Stahl and Alexeev (1985), Dye and Antle (1986), and Davidson et al. (2007)). The clandestine nature of black markets however creates transaction costs for market participants, such as potential legal consequences and search costs, which impede market performance. Whether one is interested in evaluating the efficacy of enforcement or welfare effects, an understanding of the incentives and impediments to trade is required.

In this paper, we measure the performance of the black market for license plates in Beijing. Cars driving in Beijing are required to have Beijing license plates. In a regulatory response to increasing pollution and congestion, the Beijing government rationed the number of non-transferable license plates, and hence the sales of cars, by lottery, starting January 2011. Soon thereafter, a black market for license plates emerged. We estimate the volume of trade in this market along with transaction costs and prices using comprehensive car registration data.

There is a large literature on inferring the size of the informal economy in the field of public finance, where interest tends to center on the scale of tax evasion on a macro scale. Various indirect performance measures have been proposed, from monitoring excessive electricity consumption to currency velocity, see e.g. Schneider et al. (2010). These approaches may help to infer the size of the informal economy at a macro level, they are helpful for the study of black markets for particular goods. Moreover, concerns about the literature's reliance on implausible assumptions, its weak theoretic foundations, and unreliable estimates persist (Feige (2016), Thomas (1999)). We therefore deviate from this literature and develop a simple and transparent empirical approach that is tailored to our market of interest.

The analysis proceeds in three steps. We first use standard event study methods, i.e. difference-in-difference regressions, to document a material shift in the sales of new cars towards more expensive car models. This shift is hard to explain with a lottery that randomly allocates the license plates to the car-buying households. We find no such shifts in the car sales in comparable, nearby cities where no such rationing takes place, nor do we find evidence of the shift being caused by supply side price responses to the rationing of car sales. The shift in sales towards more expensive cars is consistent with a black market that allocates the license plates to wealthier households that buy more expensive models. Similar shifts are documented in other Chinese cities that use auctions to allocate the license plates (Li and Xiao et al.). We also document price jumps in Tianjin following the introduction of a partly market based license plate allocation mechanism. The price shift by itself does however not tell us much about the volume of illegal trade.

In the next step, we combine institutional features with minimal assumptions on preferences to infer the volume of black market trade using optimal transport methods. Optimal transport has recently received renewed attention in the econometric theory literature, see e.g. Galichon (2016) for an overview, but has rarely been used in the applied literature, with the notable exception of matching models (Chiappori and Salani (2016)). Unlike the difference-in-difference analysis, which only uses conditional means of the sales distribution, optimal transport extracts information from the full sales distribution in an intuitively appealing way. That allows us to transparently estimate a lower bound on the volume of trade in the black market under weak and plausible assumptions.

In the final step, we combine our estimates of the size of the black market with estimates from Li's study of the Beijing license plate lottery to infer the black market gains from trade and transaction costs. By a Coase Theorem argument, we consider transaction costs as a tax that restricts the volume of black market trade. Following a discussion of various components of the unobserved transaction costs, we infer these costs along with transaction prices from a simple market equilibrium model as those that rationalize the estimated volume of black market trade.

Compared to a market where non-transferability is strictly enforced, we find that the black market creates sizable gains from trade by reallocating licenses to households with higher willingness-to-pay. We estimate a lower bound to the allocative efficiency gain of RMB 9.82 billion ($1.64 billion) compared to a market where non-transferability of the license plates is strictly enforced. We find that up to RMB 6.45 billion ($1.07 billion) are lost to transaction costs.

Our paper makes two contributions. Firstly, we go beyond standard event study approaches and indirect methods to infer the black market performance under weak and plausible assumptions on preferences. We believe that optimal transport has broader applicability to event studies in general and the study of black markets in particular. Secondly, our study complements the literature on the impact of the recent rationing of car sales in larger Chinese cities, e.g. Li (2018), Xiao et al. (2017), and Tan et al. (2019). This literature has assumed away the existence of a black market. We show that the size of the black market is material: at least 12% of the quota is traded in these markets, with a 95% confidence interval of (6.1%,16.7%) on the lower bound. The size of the market has immediate implications for enforcement and regulation.

We give an overview of the lottery rationing mechanism and the black market in Section 2. We describe the data in Section 3 and do a standard event study in Section 4. The optimal transport methods and market size estimation results are laid out in Section 5. The black market equilibrium model is developed and evaluated in Section 6 along with the implied estimated gains from trade and transaction costs. Some discussion is offered in Section 7.

2 The license plate lottery and the black market

The Beijing government announced in late December 2010 that, effective January 2011, car license plates would be rationed to a quota of about 35% of the previous year's sales. The license plates would be allocated by lottery. A Beijing household needs a license plate before it can register a new car. The license plates are non-transferable. The application process was simple: the pecuniary costs are low and the application can be quickly completed online. Applicants have a uniform probability of winning a plate and there are limits on the number of applications each household can submit. Beijing has strict regulations that restricts the use of cars that are not registered in Beijing.

Immediately following the introduction of the lottery, the number of newly registered cars in Beijing was reduced to the quota. The quotas have since been tightened further. Though the Beijing government has not officially justified its choice of a lottery over a market based allocation mechanism, e.g. an auction, Li (2018) speculates that a lottery was chosen out of fairness concerns.[1]

Distributing the license plates by lottery creates gains from trade: rationed prospective car buyers may compensate either government officials or legitimate lottery winners for a license plate. Despite the official non-transferability of the license plates, there are numerous news reports of the emergence of a black market. The New York Times reported that the head of the Beijing department of transportation was sentenced to life in prison in 2015 for selling certain license plates that allowed traffic privileges. According to South China Morning Post, officials involved with drafting the lottery rules had unusual luck in winning in the lottery.[2] The same article reports of widespread accusations of the lottery being rigged.

Lottery winners are a second, and reportedly more material source of supply. Reuters and the New York Times report of a number of online sites that match license owners with buyers. The going yearly rental price for a license plate in 2013 was reportedly around RMB 5000, or about half the average monthly wage in Beijing. Owners of license plates may alternatively rent out their cars. Given an interest rate at the time of about 6%, a fifteen year NPV of a license is then about RMB 65 000. Other sources however report of license purchase prices of RMB 200 000, or about twice the price of the bestselling car model (Ford Focus), see South China Morning Post. Enforcement appears to be lax. In

[1] In its footnote 13, Li reports evidence from a survey in 2013 which showed that even though a majority of Beijing residents recognized the need for rationing, less than 10% preferred an auction, and about 40% preferred a hybrid lottery and auction mechanism.

[2] "The Beijing News reported on Thursday that a record 1.26 million residents competed for fewer than 20,000 plates this month. "Liu Xuemei" was so lucky that this person - or perhaps people - won two plates in May. Cynical car-plate hunters wondered if they should change their name to stand a better chance in the next contest. Eagle-eyed internet users soon discovered Liu Xuemei was the name of the director of the vehicle and driver management department of the Ministry of Public Security. Liu Xuemei, in her 30s, is in charge of drafting rules for vehicle permits."

2014, Reuters quotes a sales representative of a Peugeot dealership, with full name, who claims that his dealership can provide prospective buyers the required license plates.[3]

In sum, these reports suggest that non-transferability is leniently enforced and give some indication on the transaction prices. The evidence is however anecdotal and gives little indication as to the size of the black market.

3 Data

We obtained vehicle registration data from a Chinese marketing research company, the Webinsight Technology and Information Corporation[4]. We also collected public information on vehicle quotas and city characteristics from other channels specified below. The vehicle registration data have observations on city-month level on aggregate registration of all vehicle models available from January 2010 to December 2015 in 35 China cities. The vehicle models are identified by their unique codes as introduced by Motor Vehicles' Type and Model Designation, published by the National Standard of People's Republic of China, and various car characteristics are provided. Table 2 presents the summary statistics.

The vehicle quota data is collected from each city's official publications.[5] The data includes the number of applicants, the quota, the average value of winning bids (in the auction rationing mechanism), and the share of each type of rationing mechanism if multiple mechanisms are applicable (see Table 1). Table 2 aggregates these variables across cities and years.

Table 1: Rationing mechanisms

City	2015 Per Capita GDP (USD)	Rationing Mechanism	Announcement Date	Implementation Date	Average Quota per Month	Quota Allocation (electric:lottery:auction)	Restriction on vehicles without local plate
Beijing	18731	Lottery	12/23/2010	1/1/2011	10,000	N/A	Yes
Tianjin	13355	Auction and Lottery	12/15/2013	12/16/2013	8,000	1:5:4	Yes
Shijiazhuang	4576	No rationing	n/a	n/a	n/a	n/a	n/a

[3] "Wang Shaoyong, sales manager at a Peugeot dealer in Beijing, said his shop provides car buyers with license permits from a partner firm that has many car plates registered in its name."

[4] http://www.webinsight.cn

[5] The official publications of information on auctionlottery are from Shanghai International Commodity Auction Co. Ltd., Beijing Passenger Vehicle Quota Administration Office, Tianjin Information System of Passenger Vehicle Quota Administration

Table 2: Summary statistics

Variables	count	mean	std dev	min	max
MSRP (RMB)	5 065 520	150 356	100 919	20 800	1 305 000

4 Event study

In Section 2, we documented widespread reports of a black market for license trades. Figure 1 documents a 20% jump in the average sales price of new cars following the introduction of the lottery. It is hard to see how the Beijing lottery, with a uniform probability of winning, would shift the sales distribution if non-transferability of licenses was strictly enforced. If preferences remained unchanged, the lottery would presumably draw winners from the same distribution of preferences. We would then expect the sales distribution would to remain unchanged after the lottery introduction.

The shift in the average prices may be due to economic trends unrelated to the rationing. We therefore compare the Beijing prices to those in the cities of Tianjin and Shijiazhuang, which are 130 km and 270 km away from Beijing, respectively. Tianjin is the fourth largest city in China. It was one of fastest growing cities in the period covered by the data, with about $\frac{2}{3}$'s of the GDP of Beijing. Shijiazhuang is the capital and the largest city of North China's Hebei Province. It has about 23% of the GDP of Beijing, but a similar population size. Neither Tianjin or Shijiazhuang rationed the sales of new cars before 2014. Figure 1 shows that while the average price of a registered car increased by 25% in Beijing from 2010 to 2011, the average price increases in Tianjin and Shijiazhuang were much more modest in the same period.

Tianjin introduced a rationing mechanism that allocates licenses using both lottery and auctions in 2014. There, 50% of the licenses are allocated by lottery, 40% are allocated by auction, and the remainder is for electric cars. The auction mechanism presumably allocates its share of the licenses towards richer households, while there may or may not have emerged a black market for the lottery allocated licenses. Using an auction expectedly changes the sales distribution towards consumers with higher higher incomes and willingness to pay. Consumers with higher incomes tend to bid more for a license plate, are therefore more likely to win, and subsequently purchase more expensive cars, which is verified empirically by Tan et al. (2019) and Xiao et al. (2017). Figure 1 shows a price jump in Tianjin following the introduction of the partly market based rationing mechanism in 2014 that is similar to, though more modest than, the one in Beijing in 2011. The Tianjin price jump suggests that a similar (black) market mechanism also reallocates the license plates to richer consumers in Beijing.

Price jumps are observed in other Chinese cities using market based mechanisms to ration car sales. Shanghai has used an auction to allocate license plates since 1996. The Shanghai auction prices have risen steadily, from about 20% of the average car price in 2010 to about 40% in 2015.[6]

The jump in the average car price in Beijing could be due to car dealers adjusting their pricing in response to regulation that reduced the demand by about half. Figure 2 shows the distribution of the change in MSRP for car models that were registered in both 2010 and 2011. It is clear that there were hardly any changes in the car model prices. The lack of supply side pricing responses is probably explained by vertical contracts that were common in the Chinese car industry at the time. Li reports that manufacturers used resale price maintenance at the national level, which precludes car dealerships from adapting their pricing to contractions in local demand.

We run the following difference-in-difference specifications in logs for three control groups: Tianjin, Shijiazhuang, and Tianjin and Shijiazhuang combined.

$$p_{j,c,t} = \alpha_0 + \alpha_1 Beijing_{j,c,t} + \alpha_2 post_{j,c,t} + \alpha_3 Beiijng_{j,c,t} \times post_{j,c,t} + \epsilon_{j,c,t} \quad (1)$$

where $p_{j,c,t}$ is the jth price observation in city c, in month t, and where $Beijing_{j,c,t}$ is an indicator which is one for price observations from Beijing, and zero for cities in the control group. The indicator $post_{j,c,t}$ is one in all months after the introduction of the lottery in Beijing, zero otherwise. Table 3 shows that the estimated price jump, $\hat{\alpha}_3$, are similar and around 17% across the specifications.

We conclude that the shift in the average prices is consistent with a black market

[6] The average Shanghai auction price in 2010 was RMB 39 000 and increased to RMB 81 000 in 2015, while the average car price increased from RMB 173 000 in 2010 to RMB 215 000 in 2015.

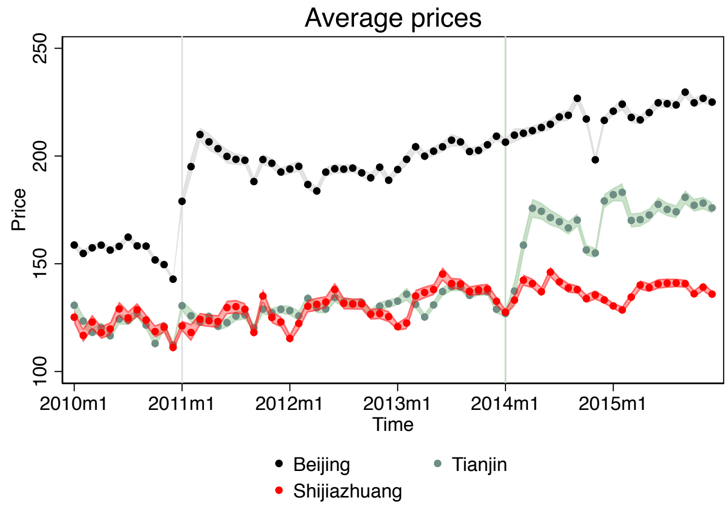

Figure 1: Tianjin introduced a hybrid lottery and auction mechanism January 2014. Confidence bands at 0.1 percent level.

for license plates. It is however hard to infer the volume of trade from shifts in the average prices, which is only one summary statistic of the shifts in the sales distributions. In the next section, we use shifts in the full sales distributions to derive an estimate of the volume of illegal trade.

5 Estimating the size of the black market

Figure 3 shows overlaid, smoothed empirical distributions of car prices in Beijing in 2010 and 2011. The sales distribution clearly shifts to the right post-lottery, which is unexpected if preferences for cars are stable over time, independent of winning, and non-transferability of license plates is enforced. Changes to the sales distribution is however consistent with a black market that reallocates the licenses to households that buy more expensive cars. Our goal is to find a credible estimate of the share of unobserved trades under minimal assumptions. An identification problem is that we do not directly observe license plate trades, only car purchases.

Our strategy is to infer a lower bound for the share of trades in the black market by quantifying the displacement of the car sales distributions in Figure 3. We show that a straightforward and transparent way of estimating a lower bound for the number of unobserved trades is to cast it as an optimal transport problem. Optimal transport has a

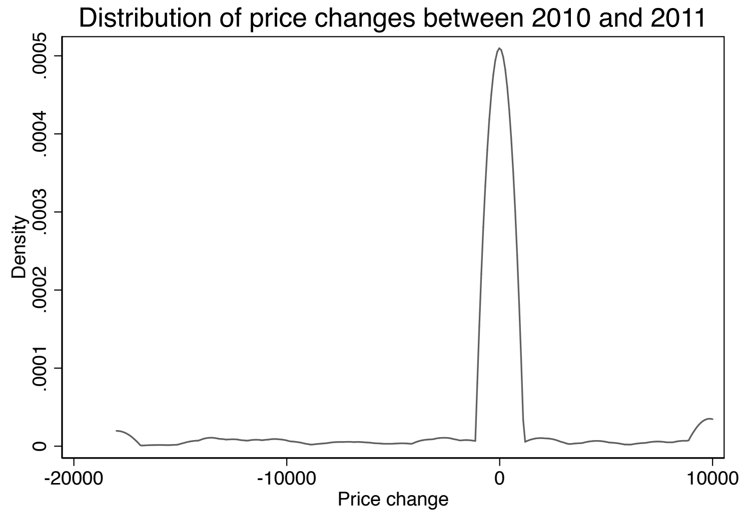

Figure 2: Distribution of price changes

long history in economics and operations research, see Kantorovitch (1958) for an early treatment. Optimal transport has recently witnessed renewed interest in economics and applied econometrics. It has found sophisticated uses in both economic and econometric theory, for instance in the analysis of identification of dynamic discrete-choice models (Chiong et al. (2016)), in vector quantile regression (Carlier et al. (2016)), and in empirical matching models (Galichon et al. (2018)).

A point we press here is that optimal transport is also a natural and easily implementable method for transparent, applied empirical analysis under minimal and economically motivated assumptions. As in standard regression analysis, we must choose a specification. When carrying out standard regression analysis, the analyst selects a group of covariates and perhaps some transformations of those. When carrying out applied optimal transport analysis, the analyst selects a cost matrix describing the cost of transport, or distance, between points in the support of the compared probability mass functions. Our application below serves as an illustration. We also show that standard errors obtain and regression output may be presented in a way that is typical of regression tables.

We use two estimators. The first relies on a before-and-after comparison of the sales distributions in Beijing. The second estimator, which we develop in the next section, uses observable changes in the sales distribution in Tianjin to control for external common

Table 3: Diff-in-diff regressions for different control groups

	(1) Tianjin	(2) Shijiazhuang	(3) All
Beijing	0.266***	0.233***	0.0829***
	(172.32)	(116.76)	(74.23)
post	0.0572***	0.0438***	0.0618***
	(35.30)	(19.97)	(61.75)
Beijingxpost	0.169***	0.182***	0.165***
	(88.06)	(75.52)	(114.54)
Constant	-2.322***	-2.288***	-2.139***
	(-1713.50)	(-1233.68)	(-2563.36)
r2	0.08	0.07	0.03
N	2116640	1795962	3214675

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

trends. For the first estimator, we make the following two assumptions.

Assumption 1 *Car dealerships did not change the pricing in the period covered by the data.*

Assumption 1 is largely verified in the data, see Figure 2, and in line with the resale price maintenance contracts that the industry used at the time (Li).

Assumption 2 *The demand of prospective car buyers did not change within the period covered by the data.*

Assumption 2 is untestable. It implies that a household that wins the lottery would purchase the same car as in a world without rationing. It rules out trends in car preferences, e.g. increasing demand for SUVs, and it rules out general equilibrium effects and externalities. For instance, rationing may lead to less congestion, which increases the willingness to pay for cars. Since the rationing controls the influx of new cars to Beijing (on the order of 250 000), and not the stock (on the order of five million), we believe that such general equilibrium effects are of second-order.

Though Assumption 2 is common in the literature on rationing, a literature that was particularly active post WWII, it has been contested in the past. One example is Tobin (1952) which caters the idea that rationing can change tastes over time:

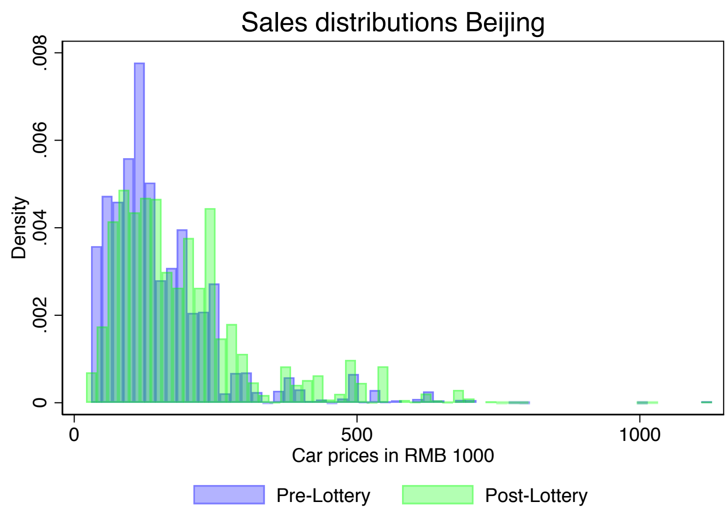

Figure 3

> "*Experience under rationing may alter the consumer's scale of preferences. He may learn to like pattern of expenditures into which rationing forces him, or to dislike it even more intensely that if it had not been forced upon him.*" (p. 548)

It is unclear how to empirically test an effect of the lottery itself on preferences against our own assumption of stable preferences and black market trade with the data we have. Such effects on preferences are ruled out by Assumption 2.

In the following, we refer to a *buyer* as a household that purchases a car with a license plate it has bought or rented illegally. We refer to a *seller* as a household that acquired a license plate and then either decided to sell or rent it to a buyer. We refer to the transaction between a buyer and a seller as a *trade*. Let $\mathbb{P}_0$ be the probability distribution that sales are drawn from pre-lottery, and $\mathbb{P}_1$ the analogous distribution post-lottery. If Assumption 1 and 2 hold, and the lottery is truly selecting buyers at random from the distribution of prospective car buyers, the sales distributions should not change from the pre-to the post lottery period, i.e. $\mathbb{P}_0 = \mathbb{P}_1$.

If license plates are instead reallocated by black market trade, then the observed sales distributions may shift over time such that $\mathbb{P}_0 \neq \mathbb{P}_1$. It seems plausible that richer buyers buy license plates from poorer sellers. Previous literature has shown that richer buyers buy more expensive cars (Li (2018), Xiao et al. (2017), and Tan et al.

(2019)), which causes a shift in the sales distribution. Phrased directly in terms of the probability distributions, trade implies that some probability mass must be moved from the pre-lottery sales distribution $\mathbb{P}_0$ to equate it to the post-lottery sales distribution $\mathbb{P}_1$. It is immediate that the minimum amount of mass required to transform $\mathbb{P}_0$ into $\mathbb{P}_1$ bounds the true mass from below, that is, it lower bounds the volume of black market trade.

If we observed $\mathbb{P}_0$ and $\mathbb{P}_1$ directly, we could simply compute the desired lower bound via the following oracle problem

$$\begin{aligned} OT(\mathbb{P}_0, \mathbb{P}_1) = \min_{\gamma \in \Gamma} & \int c(x_1, x_2)\gamma(x_1, x_2)dx_1 dx_2 \\ & \text{subject to} \\ & \int \gamma(x_1, x_2)dx_2 = \mathbb{P}_0(x_1), \quad \text{for all} \quad x_1, \\ & \int \gamma(x_1, x_2)dx_1 = \mathbb{P}_1(x_2), \quad \text{for all} \quad x_2, \end{aligned} \tag{2}$$

where $x_1, x_2 \in \mathcal{X}$ are points in the price support $\mathcal{X}$, and where $\Gamma = \{\gamma : \gamma \geq 0, \int \gamma(x_1, x_2)dx_1 dx_2 = 1\}$ is the set of all bivariate probability distributions on $\mathcal{X}$. The problem asks us to search for a bivariate probability distribution that minimizes a cost of transport $c(x_1, x_2)$ between x_1 and x_2, such that its marginal distributions are exactly equal to the observed sales distributions $\mathbb{P}_0(x)$ and $\mathbb{P}_1(x)$. We pick the cost function

$$c(x_1, x_2) = \mathbb{1}(|x_1 - x_2| > 0). \tag{3}$$

which assigns the same cost to any move between $x_1 \neq x_2$, and zero cost to $x_1 = x_2$. The optimal transport cost for the problem in (2) with this cost function can be interpreted as the smallest share of illegal trades necessary to rationalize a sales distribution $\mathbb{P}_1$ given $\mathbb{P}_0$.

Our estimand is then

$$s(\mathbb{P}_0, \mathbb{P}_1) = OT(\mathbb{P}_0, \mathbb{P}_1). \tag{4}$$

An example may be instructive. Suppose cars are sold either at a price of 1 or a price of 2, so $\mathcal{X} = \{1, 2\}$. Suppose four cars are sold pre-lottery, three at price 1 and one at price 2, such that $\mathbb{P}_0 = \frac{1}{4}[3, 1]$. Suppose four cars are sold post-lottery, one at price 1 and three at price 2, such that $\mathbb{P}_1 = \frac{1}{4}[1, 3]$. Assuming no sampling variation, it is immediately clear that at least two license plates must be traded to rationalize the shift in the sales distribution. One solution γ which satisfies the constraints (2) is

$$\gamma = \frac{1}{4}\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}.$$

The optimal transport is calculated by summing all off-diagonal terms of γ. The diagonal terms imply no transport, and hence no cost. We get $OT(\mathbb{P}_0, \mathbb{P}_1) = 50\%$ as a lower bound of the share of illegal trade.

There may be more than one γ that solves (2). Our parameter of interest is however not γ, but the transport cost $OT(\mathbb{P}_0, \mathbb{P}_1)$, which is clearly unique. Multiplicity of solutions γ therefore has no implications for identification.

Though richer buyers purchase more expensive models, it is not obvious that trade in license plates shifts the sales distribution towards more expensive car models. The transaction price for a license plate serves as a tax on the car purchase which may lead a richer buyer to trade down relative to the car that the same buyer would have purchased if she did not have to buy a license plate first. The estimand in (4) however does not require that buyers purchase more expensive cars than the sellers would have purchased if they could not trade, i.e., that the sales distribution shifts to the right. We get a lower bound to the share of trade as long as the sellers would have purchased different cars than the buyers purchase.

Since all license plates could be traded without shifting the distribution, we can not derive a useful upper bound by comparing the sales distributions. This would for instance be the case if every seller traded with a buyer at its point in the sales distribution for a negligible transaction prices. Without further assumptions, the upper bound to the share of trades is therefore one. In Section 6, we however show that under further assumptions on who trades in the black market, the data bound the trade from above to 61% from the data.

We next need to account for the fact that we have the sample distributions, and not the population distributions. The sample distribution pre-lottery is $\hat{\mathbb{P}}_{0,n_0}$, where n_0 is the number of observations, and $\hat{\mathbb{P}}_{1,n_1}$ and n_1 is defined analogously for the post-lottery period. We approximate the population problem in (2) with its discrete sample equivalent

$$
\begin{aligned}
OT(\hat{\mathbb{P}}_{0,n_0}, \hat{\mathbb{P}}_{1,n_1}) = & \min_{\gamma \in \Gamma} \sum_{i,j} \gamma_{i,j} \mathbf{C}_{i,j} \\
& \text{subject to} \\
& \sum_{i} \gamma_{i,j} = \hat{\mathbb{P}}_{0,n_0}(j), \quad \text{for all} \quad j, \\
& \sum_{j} \gamma_{i,j} = \hat{\mathbb{P}}_{1,n_1}(i), \quad \text{for all} \quad i, \\
& \gamma_{i,j} \geq 0 \ \text{ for all } i, j.
\end{aligned} \tag{5}
$$

where $\mathbf{C} = \mathbf{1}\mathbf{1}^T - \text{diag}(\mathbf{1})$ interacted with γ. However, this program may confound sampling uncertainty for trades. We next discuss how to choose $\mathbf{C}$ to minimize the cost of moving the distribution from $\hat{\mathbb{P}}_{0,n_0}$ to $\hat{\mathbb{P}}_{1,n_1}$ for our applied optimal transport analysis, while taking sampling uncertainty into account.[7]

5.1 Choosing tuning parameters using a placebo test

The observed sample distributions $\hat{\mathbb{P}}_{0,n_0}$ and $\hat{\mathbb{P}}_{1,n_1}$ are plotted in Figure 4. We cannot directly apply the program in (5) to $\hat{\mathbb{P}}_{0,n_0}$ and $\hat{\mathbb{P}}_{1,n_1}$ due to sampling uncertainty. Even in the null case of no trade in license plates, i.e., if both $\hat{\mathbb{P}}_{0,n_0}$ and $\hat{\mathbb{P}}_{1,n_1}$ were composed of draws from the same distribution $\mathbb{P}_0$, sampling variation would lead to differences in the realized distributions.

One approach to smooth out sampling variation is to ignore small moves: we can attribute zero cost to transport mass up to some small distance $d \geq 0$, where the distance

[7] Because $n_1 < n_0$, we can normalize such that both $\hat{\mathbb{P}}_{0,n_0}$ and $\hat{\mathbb{P}}_{0,n_1}$ sum to n_1, and thus $OT(\hat{\mathbb{P}}_{0,n_0}, \hat{\mathbb{P}}_{1,n_1})$ gives a bound on the number of trades.

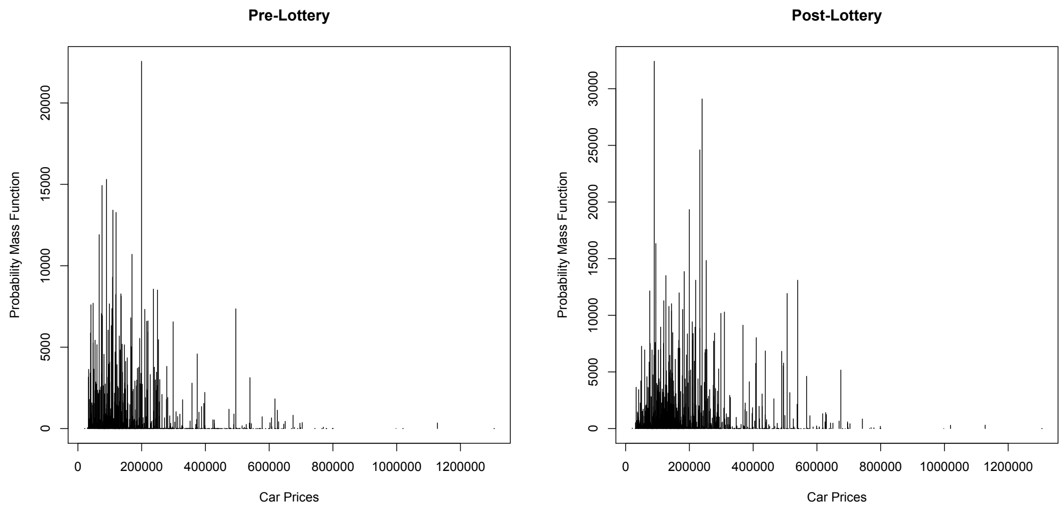

Figure 4: Exact empirical sales distributions before-and-after.

is measured in the purchase price of the car. To implement this smoothing criterion, we construct a cost matrix $\mathbf{C}(d)$ with entries

$$\mathbf{C}_{i,j}(d) = \mathbb{1}\left(|x_i - x_j| > d\right),$$

where x_i is the i^{th} entry of $\mathcal{X}$. We then solve the sample optimal transport problem

$$\begin{aligned} OT_d(\hat{\mathbb{P}}_{0,n_0}, \hat{\mathbb{P}}_{1,n_1}) = \min_{\gamma \in \Gamma} & \sum_{i,j} \gamma_{i,j} \mathbf{C}_{i,j}(d) \\ & \text{subject to} \\ & \sum_i \gamma_{i,j} = \hat{\mathbb{P}}_{0,n_0}(j), \text{ for all } j, \\ & \sum_j \gamma_{i,j} = \hat{\mathbb{P}}_{1,n_1}(i), \text{ for all } i, \\ & \gamma_{i,j} \geq 0 \text{ for all } i, j. \end{aligned} \tag{6}$$

We choose the parameter d such that the estimated transport cost between two distributions sampled from $\mathbb{P}_0$ is approximately zero, that is, when we know that the cost in the population problem is exactly zero. In order to calibrate d on pairs of empirical distributions drawn from $\mathbb{P}_0$, we use $\hat{\mathbb{P}}_{0,n_0}$ as an estimate of $\mathbb{P}_0$. We then obtain new simulated empirical distributions by drawing $X_{j,1}, ..., X_{j,n_0}$ from $\hat{\mathbb{P}}_{0,n_0}$, for $j = 1, 2$, and by collecting their corresponding empirical distributions $\tilde{\mathbb{P}}^{(j)}_{0,n_0}$. For a particular set of

	$d = 30\ 000$	$d = 50\ 000$	$d = 70\ 000$	$d = 90\ 000$
$\hat{s}(d)$	12%	7%	6%	5%
$\hat{s}(d)_{placebo}$	0.03%	0.001%	0.001%	0.001%

Table 4: Relative transport costs.

simulated distributions, the placebo transport costs are $OT_d(\tilde{\mathbb{P}}^{(1)}_{0,n_0}, \tilde{\mathbb{P}}^{(2)}_{0,n_0})$ from program (6), with marginal constraints replaced with $\tilde{\mathbb{P}}^{(1)}_{0,n_0}$ and $\tilde{\mathbb{P}}^{(2)}_{0,n_0}$. Our estimator of the placebo transport costs is $\hat{s}_{placebo}(d) = \mathbb{E}\left[OT_d(\tilde{\mathbb{P}}^{(1)}_{0,n_0}, \tilde{\mathbb{P}}^{(2)}_{0,n_0})\right]$, where the average is over the set of simulated distributions.

We next probe the sensitivity of the estimates to a range of d's. Figure 6 plots $\hat{s}(d) = OT_d(\hat{\mathbb{P}}_{0,n_0}, \hat{\mathbb{P}}_{1,n_1})$ and $\hat{s}_{placebo}(d)$ against d. While $\hat{s}(d)$ estimates the lower bound to the market share, $\hat{s}_{placebo}(d)$ estimates the sampling uncertainty as a function of d. A larger value of d reduces sampling variation, while a smaller value of d reduces bias by exploiting more of the information in the sample. We want d as small as possible, while maintaining negligible sampling variation. We select d such that the transport cost due to sampling variation in the placebo problem is at least an order of magnitude smaller than the estimated transport cost for the target problem.

The results in Table 4 suggest that using a d larger than 30 000 would be overly conservative. The placebo transport cost $\hat{s}_{placebo}(30\,000)$ is already small in absolute terms and multiple orders of magnitude smaller than $\hat{s}(30\,000)$. This tells us that sampling uncertainty is negligible at this d, so we choose $\hat{s}(30\,000) = 12\%$ as our preferred estimate.

5.2 Difference-in-differences lower bounds

The results in the previous section relied on an assumption of a stationary sales distribution in Beijing, but for the lottery (Assumption 2). This assumption may not hold. For instance, income growth or changes in car preferences could lead the sales distributions to shift for reasons that are unrelated to black market trade. In Section 4, we used the sales data from Tianjin to control for common trends using difference-in-difference regressions.

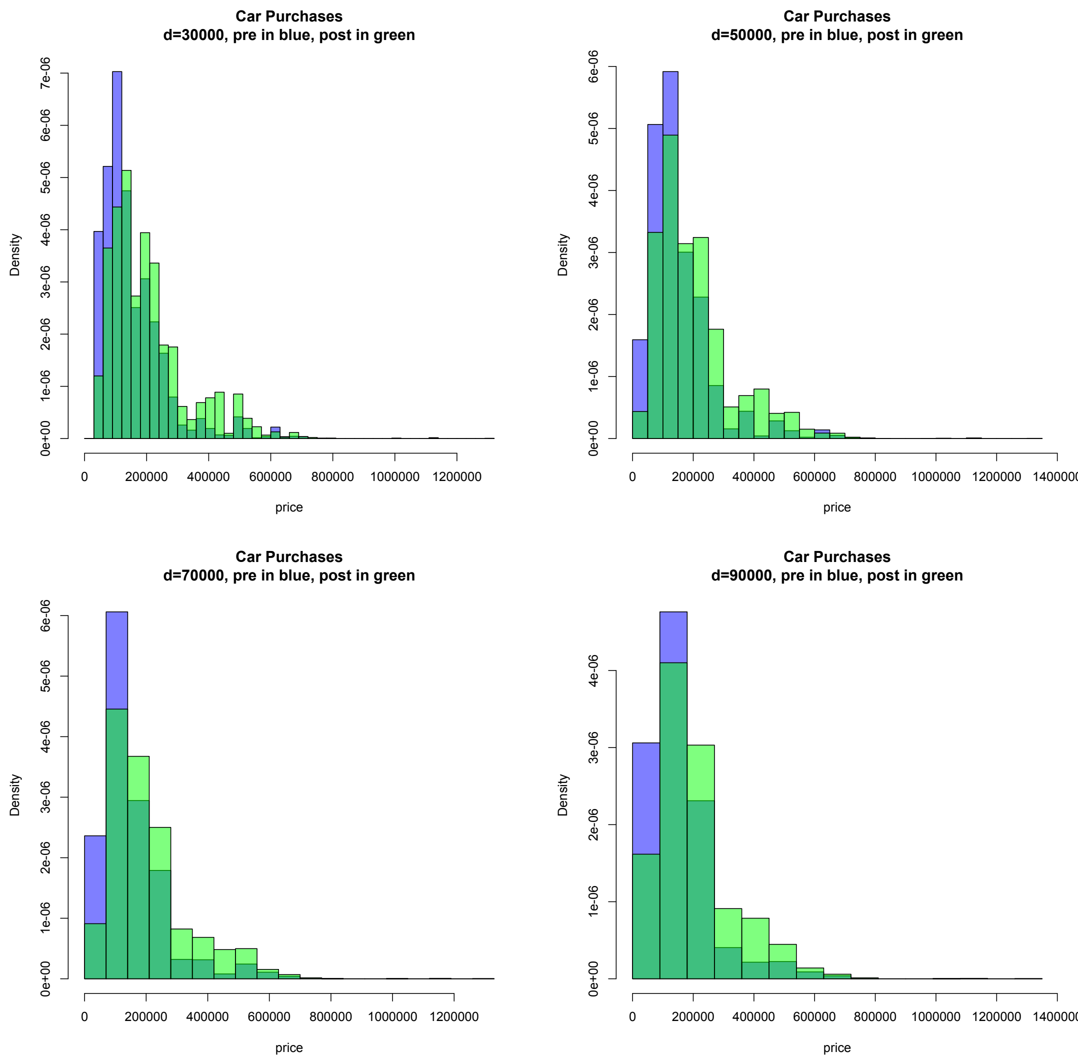

Figure 5: Distributions smoothed at all candidates d's.

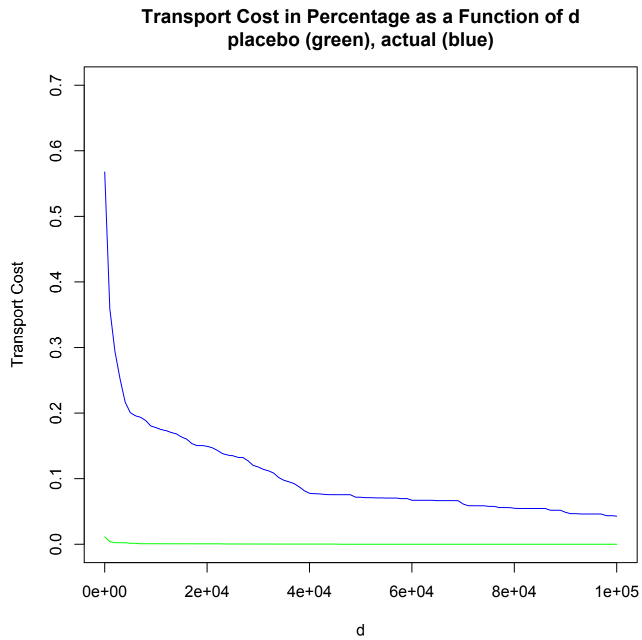

Figure 6: True and placebo costs as a function of d

We draw on the same idea for our second black market share estimator. Figure 7 shows that the sales distribution in Tianjin, where there was no rationing, arguably shifted to the right, but less pronouncedly than the shift in Beijing in the same period, as seen in Figure 3. We would like to use this observable shift in the sales distribution in Tianjin to account for external factors that would have shifted the sales distribution in Beijing, had there not been a lottery, under the assumption that both cities were exposed to common trends.

It is not immediately clear what the analogy to the common trends assumption in standard difference-in-difference analysis is for optimal transport. While a common trend is about a direction, optimal transport is about distance, which by definition has no direction. We make two assumptions below to capture an idea analogous to the common trends assumption. The first assumption restricts the mass displaced by external factors to be of the same volume in Beijing and Tianjin. The second assumption restricts that mass to shift in a particular direction (in our case, to the right). Taken together, we say that these two assumptions make up the *common displacement*, which is our analogue to the concept of common trends for distributions.

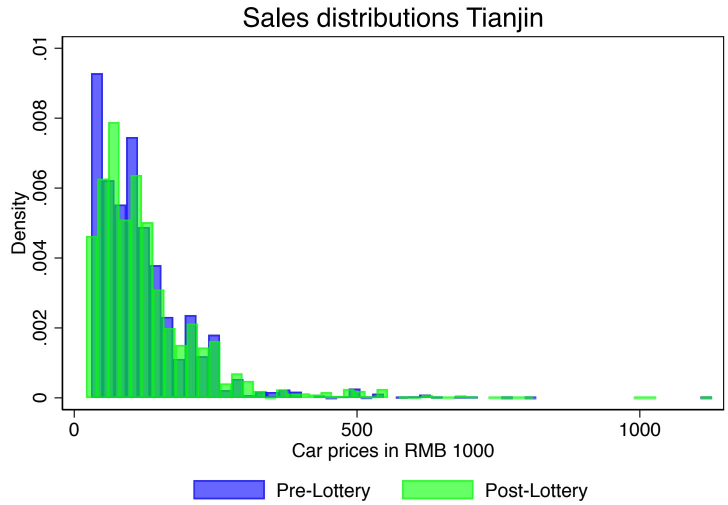

Figure 7

It is useful to formalize the idea in potential outcomes notation. Let $\mathbb{P}_{Beijing,pre}(t)$, $\mathbb{P}_{Tianjin,pre}(t)$, $\mathbb{P}_{Beijing,post}(t)$, and $\mathbb{P}_{Tianjin,post}(t)$ be the sales distributions in Beijing and Tianjin before and after the introduction of the lottery in Beijing. Let $t = 1$ index the potential outcome distributions realized under the lottery policy, and $t = 0$ index those under the status quo of no lottery.

The target quantity is

$$OT\left(\mathbb{P}_{Beijing,post}(1), \mathbb{P}_{Beijing,post}(0)\right), \tag{7}$$

the displacement between the observed Beijing sales distribution in the post lottery period and the counterfactual Beijing sales distribution in the post lottery period, had there not been a lottery. Its oracle difference-in-differences proxy is given by the transport cost for the observed Beijing sales distributions before and after the lottery, net of the transport cost that would have been incurred if there was no Beijing lottery. We write this quantity as

$$s_{did} = OT\left(\mathbb{P}_{Beijing,post}(1), \mathbb{P}_{Beijing,pre}(0)\right) - OT\left(\mathbb{P}_{Beijing,post}(0), \mathbb{P}_{Beijing,pre}(0)\right). \tag{8}$$

By the triangle inequality of Proposition 55 in Peyré (2018), the population quantity s_{did} with $\mathbf{C} = \mathbf{C}(0)$ produces a lower bound on the quantity of interest,

$$\begin{aligned} & OT\left(\mathbb{P}_{Beijing,post}(1), \mathbb{P}_{Beijing,pre}(0)\right) - OT\left(\mathbb{P}_{Beijing,post}(0), \mathbb{P}_{Beijing,pre}(0)\right) \\ & \leq OT\left(\mathbb{P}_{Beijing,post}(1), \mathbb{P}_{Beijing,post}(0)\right). \end{aligned} \tag{9}$$

This parameter depends on the unobserved quantity $OT\left(\mathbb{P}_{Beijing,post}(0), \mathbb{P}_{Beijing,pre}(0)\right)$. Under a common trends like assumption that we define below, we may recuperate the analog of the traditional difference-in-differences equation, which is an identifiable lower bound. We relax Assumption 2 and substitute it with the following *equal displacement* assumption.

Assumption 3 *The optimal transport distance between the sales distributions before and after the Beijing lottery, but for the lottery, is the same in Beijing and Tianjin. In potential outcomes notation:*

$$OT\left(\mathbb{P}_{Tianjin,post}(0), \mathbb{P}_{Tianjin,pre}(0)\right) = OT\left(\mathbb{P}_{Beijing,post}(0), \mathbb{P}_{Beijing,pre}(0)\right). \tag{10}$$

Under Assumption 3, the oracle difference-in-differences estimator in the population in (8) can be written in terms of observable optimal transport distances, delivering the optimal transport analog of the traditional difference-in-differences estimator,

$$s_{did} = OT\left(\mathbb{P}_{Beijing,post}(1), \mathbb{P}_{Beijing,pre}(0)\right) - OT\left(\mathbb{P}_{Tianjin,post}(0), \mathbb{P}_{Tianjin,pre}(0)\right). \tag{11}$$

Although s_{did}, under Assumption 3, is well estimated by

$$OT_d\left(\hat{\mathbb{P}}_{Beijing,post}(1), \hat{\mathbb{P}}_{Beijing,pre}(0)\right) - OT_d\left(\hat{\mathbb{P}}_{Tianjin,post}(0), \hat{\mathbb{P}}_{Tianjin,pre}(0)\right)$$

for a small, but non-zero value of d, we would like our difference-in-differences estimate to provide a lower bound to the sample version of the target quantity, i.e. $OT_d\left(\hat{\mathbb{P}}_{Beijing,post}(1), \hat{\mathbb{P}}_{Beijing,post}(0)\right)$. We show below that the following estimator delivers this property.

$$\hat{s}_{did}(d) = OT_{2d}\left(\hat{\mathbb{P}}_{Beijing,post}(1), \hat{\mathbb{P}}_{Beijing,pre}(0)\right) - OT_d\left(\hat{\mathbb{P}}_{Tianjin,post}(0), \hat{\mathbb{P}}_{Tianjin,pre}(0)\right). \tag{12}$$

Under Assumption 3, and for d sufficiently large, $\hat{s}_{did}(d)$ is a proxy for

$$OT_{2d}\left(\hat{\mathbb{P}}_{Beijing,post}(1), \hat{\mathbb{P}}_{Beijing,pre}(0)\right) - OT_d\left(\hat{\mathbb{P}}_{Beijing,post}(0), \hat{\mathbb{P}}_{Beijing,pre}(0)\right), \tag{13}$$

which we can show lower bounds the sample equivalent of the population quantity of interest.

Theorem 1 *Let $OT_d\left(\hat{\mathbb{P}}, \tilde{\mathbb{P}}\right)$ be the discrete optimal transport problem described in (6) with $\mathbf{C} = \mathbf{C}(d)$, let $\hat{\mathbb{P}}_a$, $\hat{\mathbb{P}}_b$ and $\hat{\mathbb{P}}_c$ be three probability mass functions. Then, the following inequality holds*

$$OT_{2d}\left(\hat{\mathbb{P}}_a, \hat{\mathbb{P}}_b\right) - OT_d\left(\hat{\mathbb{P}}_c, \hat{\mathbb{P}}_b\right) \leq OT_d\left(\hat{\mathbb{P}}_a, \hat{\mathbb{P}}_c\right). \tag{14}$$

Proof: This follows immediately from the argument of Proposition 55 in Peyré (2018), and the analogous triangle inequality for the pseudo-distances $\mathbb{1}\left(|\chi_i - \chi_j| > d\right)$, where χ_i is the i^{th} entry of χ, the discrete support of car prices. □

In particular, Theorem 1 stipulates that if the equal displacement assumption holds in sample, i.e., $OT_d\left(\hat{\mathbb{P}}_{Beijing,post}(1), \hat{\mathbb{P}}_{Beijing,post}(0)\right) = OT_d\left(\hat{\mathbb{P}}_{Tianjin,post}(1), \hat{\mathbb{P}}_{Tianjin,post}(0)\right)$, then

$$\hat{s}_{did}(d) \leq OT_d\left(\hat{\mathbb{P}}_{Beijing,post}(1), \hat{\mathbb{P}}_{Beijing,post}(0)\right). \tag{15}$$

This is a desirable property: our estimator satisfies (15), the sample analog to bound (8), the quantity of interest in population.

A lower bound on $OT_d\left(\hat{\mathbb{P}}_{Beijing,post}(1), \hat{\mathbb{P}}_{Beijing,post}(0)\right)$ is useful, but it remains practical to get a sense of how tight the obtained inequality is, i.e., of how close it is to being an equality. We give a sufficient condition, tailored to the application at hand, which delivers equality. General conditions and characterization of the tightness of the relationship are examined in the Appendix.

Assumption 4 *The following stochastic dominance relationship holds between distributions,*

$$\mathbb{P}_{Beijing,pre}(0) \preceq \mathbb{P}_{Beijing,post}(0) \preceq \mathbb{P}_{Beijing,post}(1). \tag{16}$$

While Assumption 3 restricts the displacement of the Beijing sales distribution in the counterfactual world of no lottery, Assumption 4 pertains to shift the location of the distributions, analogously to the common trend assumption in standard difference-in-difference analysis. We call the two assumptions combined for *common displacement*. Under common displacement, it immediately follows that

$$\begin{aligned} &OT(\mathbb{P}_{Beijing,post}(1), \mathbb{P}_{Beijing,post}(0)) \\ &= OT(\mathbb{P}_{Beijing,post}(1), \mathbb{P}_{Beijing,post}(0)) - OT(\mathbb{P}_{Beijing,post}(0), \mathbb{P}_{Bejing,pre}(0)), \end{aligned} \tag{17}$$

and we may consider $\hat{s}_{did}(d)$ as a point estimate for, and not an estimated lower bound on, the quantity of interest $OT\left(\mathbb{P}_{Beijing,post}(1), \mathbb{P}_{Beijing,post}(0)\right)$.

The plausibility of the first relationship of Assumption 4, $\mathbb{P}_{Beijing,pre}(0) \preceq \mathbb{P}_{Beijing,post}(0)$, can be assessed by inspecting Figure 7, which plots the distributions making up the analogous difference, $\mathbb{P}_{Tianjin,pre}(0)$ and $\mathbb{P}_{Tianjin,post}(0)$, or by inspecting or testing the solution $\hat{\gamma}$ to $OT_d\left(\mathbb{P}_{Tianjin,pre}(0), \mathbb{P}_{Tianjin,post}(0)\right)$, as discussed in the Appendix on general optimal transport difference-in-differences estimators. Inspection of Figure 7 suggests that $\mathbb{P}_{Tianjin,post}(0)$ stochastically dominates $\mathbb{P}_{Tianjin,pre}(0)$.

The plausibility of the second relationship of Assumption 4, $\mathbb{P}_{Beijing,post}(0) \preceq \mathbb{P}_{Beijing,post}(1)$, depends on prior information about the nature of the main effect. It suffices to assume that all license plate buyer purchase more expensive cars than the sellers would have purchased, had they not sold the license. This seems plausible if richer people buy licenses from poorer people on the black market, and that the income effect from the license expenditure is not too great.

The difference-in-difference results are given in Table 5. The difference-in-differences

d	1 000	5 000	10 000	20 000	30 000
Beijing	31.1%	17.2%	14.4%	11.5%	8.3%
Tianjin	22.5%	5.8%	2.8%	1.7%	0.9%
Difference	8.6%	11.4%	11.6%	9.8%	7.4%

Table 5: The first two rows report the lower bound estimates for black market trade in Beijing and Tianjin. The last row reports the difference-in-difference estimate of s_{did}

estimates are relatively robust to the choice of d. As before, the parameter implies a bias variance trade-off. On the one hand, we want d as small as possible to detect changes in the distribution, which reduces bias. On the other hand, the smaller the d, the more sampling uncertainty the estimate absorbs. We pick $d = 5\ 000$ as the value for the tuning parameter. This value is much lower than for the before-and-after estimates in the previous section, which was $d = 30\ 000$. The difference reflects that choose the tuning parameter such that the estimates are relatively stable in a nearby region, which is a different criterion than for the first estimator. Both estimators however give very similar estimates.

The parameter d can be interpreted as an upper bound on the types of change in prices which we do not want to penalize in order to produce the lower bound on the volume of trade. Interpreting d as such suggests reinterpreting our cost function $\mathbf{C}_{i,j}(d)$ as applying $\mathbf{C}_{i,j}(0)$ after correcting for the idiosyncrasies –by say, going back to the data and setting the price of an idiosyncratic sale back to its standard price. This turns out to be of technical importance as it justifies the interpretation of our cost function as an approximate metric. The cost function $\mathbf{C}_{i,j}(d)$ is a metric only for $d = 0$, which in turn is a sufficient condition for the use of subsampling for inference (Sommerfeld and Munk (2018)). One may alternatively argue that a d of RMB 5 000 is negligibly small, as the 10th percentile is above RMB 50 000, and the median is above RMB 100 000.

At $d = 5\ 000$, the shift between the Beijing sales distribution before and after the lottery is 17.2% with standard error of 1.9%, and the shift between the sales distribution before and after the lottery in Tianjin is 5.8% with standard error of 1.9%. The difference-in-differences estimate of the proportion of trades is 11.4%, with a 95% confidence interval of (6.1,16.7).

Table 5 shows that the difference is stable even up to negligible values of d, although the transportation costs decrease in d for both Beijing and Tianjin. This fact has the comforting implication that our result is not particularly sensitive to the choice of the tuning parameter d. It suggests that the difference between two empirical distributions sampled from the same data generating process is captured and corrected by the difference-in-differences when they are not discarded by a large enough d. The low sensitivity to d also means that we can use the difference for transport costs corresponding to an arbitrarily small value of the tuning parameter d as an estimate. In that case, the implied transport cost is an approximate metric, and subsampling is approximately valid. The diff-in-diff estimate is close to, but slightly lower than the estimate that only uses the Beijing distribution. We choose 12% as our estimated lower bound to the market share in the following.

6 Inferring transaction costs and prices

Combining our estimated lower bound on the size of the black market with empirical results from the literature, we infer its transaction costs and prices in a market equilibrium model. We derive our model from a common version of the Coase Theorem: the initial allocation of goods is irrelevant in the absence of transaction costs. If the initial allocation is not Pareto efficient, as expected in our case when license plates are allocated by lottery, households will transact until the gains from trade are exhausted. A market price will form that equates the number of lottery winners willing to sell their license plates to the prospective buyers willing to buy a license. After all gains from trade are exhausted, the allocation is by construction efficient.

Transaction costs create frictions that reduce the volume of trade by taxing the gains from trade. Transaction costs may have both pecuniary and non-pecuniary components. One likely important transaction cost component is potential legal liabilities. For instance, a new car must be registered in the name of the legal owner of the license plate, and thus the owner, and not the driver, will in some cases be liable for traffic accidents. The fact that trade in license plates is illegal may also bring about non-pecuniary transaction costs. Search costs are another component: buyers and sellers need to meet in the market. The reported existence of online market places for license plates however suggests that search costs are small.[8] Potential enforcement, and perhaps moral hazard of the kind associated with leaving the official car ownership in someone else's name, are likely the key transaction cost components.

While there may be further transaction cost components, we do not attempt to distinguish between these. We infer transaction prices and transaction costs, which are both unobserved, from the market model. We bound transaction costs by recognizing them as a tax on trade necessary to rationalize our estimated volume of black market. In our model, the transaction costs are therefore a simple quantitative measure of the frictions in the black market.

We derive a demand and supply curve using Li's estimated willingness to pay for license plates. These estimates are represented by a function $v(n)$ which is given in (Figure 3, reference). A copy of Li's original Figure 4 is given in Figure 10 in the appendix.[9] In the following, we take Li's estimated willingness-to-pay for a license plate $v(n)$, with range [0,280], to be known without sampling variation. Denote the distribution of willingness-to-pay in the market by a cumulative distribution function $F(v)$ which returns the share of households in the market with a willingness to pay less than v, i.e. $F(v) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}(v(n) \leq v)$. Since $v(n)$ is strictly decreasing, we can recover F as

$$\frac{v^{-1}(v)}{N} = 1 - F(v).$$

We assume that only prospective car buyers enter the lottery.

Assumption 5 *The lottery draws q winners with equal probability. The winners' willingness to pay for a license plate is distributed with cumulative density function F.*

Assumption 5 rules out speculators: those who enter the lottery with no intention to buy a car if they win a license plate. We relax Assumption 5 and allow for speculators in Section 6.2. We set q = 280000 (the quota in 2011) and N = 710000 (the total sales of

[8] Bhave and Budish (2018) notes that online market places for ticket resale decreases transaction costs and increases the volume of resold tickets in the secondary market.

[9] Shanjun Li has generously shared his estimates with us. We made one modification to his estimates. While Li's Figure 3 shows positive willingness to pay for quantities far beyond the unrationed pre-lottery market equilibrium, we require $v(N) = 0$, that is, that the marginal willingness to pay for a license plate at the unrationed market equilibrium is zero. That $v(N) > 0$ in Li's estimates is perhaps an artefact of the logit assumption. Our results are however not very sensitive to this modification.

new cars in 2010 before the lottery). Assumption 5 implies that the willingness to pay of the $N-q$ buyers is also distributed with cumulative distribution function F. The gains from trade are created by the overlapping distributions of willingness to pay of the q lottery winners and the $N-q$ lottery losers. From this observation, we derive demand and supply curves.

We assume that there is a market transaction price p for a license plate and a transaction cost t. The transaction costs is equally borne by the buyer and the seller. A buyer is willing to buy a license plate if $v > p+t$, and a seller is willing to sell a license plate if $v \leq p-t$. We then get the following demand and supply functions

$$\begin{aligned} D(p,t) &= (N-q)(1-F(p+t)) \\ S(p,t) &= qF(p-t) \end{aligned} \tag{18}$$

Figure 8 illustrates a pair of demand and supply functions.[10] Suppose first that there are no transaction costs. Demand equals supply at price p_{notc} and quantity q_{notc}. The equilibrium sales q_{notc} are less than the quota q since lottery winners with valuation in excess of the market clearing transaction price prefer not to trade. Under the assumptions of the model, we can therefore further bound the share of trade in the illegal market from above at $s_{upper} = \frac{q_{notc}}{q} = \frac{N-q}{N} = 61\%$. The upper bound on the share of illegal trade implies that though our lower bound estimate is $\hat{s} = 12\%$ of the quota, it is $\frac{12\%}{61\%} = 20\%$ of the largest number of trades that this market can support, i.e. the number of trades that occurs when there are no transaction costs.

Suppose next that $t > 0$. The transaction costs are seen to drive a wedge between the supply and demand that lowers the volume of illegal trade to sq. The total transaction costs, summed over buyers and sellers, that rationalize the trade are hence $2tsq$.

We do not observe transaction prices and costs, but can recover these given the estimated volume of trade sq. We use $s = 12\%$ as our preferred estimate from Section 5.2.

[10] This illustration does not use the demand and supply curves that are derived from Li's estimates.

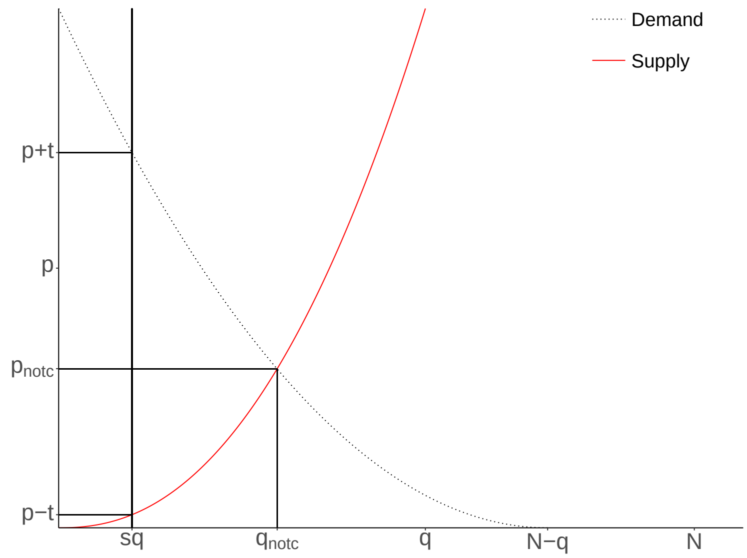

Figure 8: Illustration of demand and supply derived from $v(n)$.

Theorem 2 *Suppose that F is known and that the market clears at transaction price p and transaction cost t for a known volume of sq black market trades. Then the market clearing transaction prices and costs are identified.*

Proof:
For any pair of transaction prices and costs, the valuation of the marginal seller is $v_{seller} = p - t$ and the valuation of the marginal buyer is $v_{buyer} = p + t$. Together, we get

$$
\begin{aligned}
p &= \frac{1}{2}\left(v_{seller} + v_{buyer}\right) \\
t &= \frac{1}{2}\left(v_{buyer} - v_{seller}\right).
\end{aligned}
\tag{19}
$$

Equating demand to supply using (18) at the estimated lower bound of trades sq, the marginal valuations are uniquely recovered by inverting $F^{-1}(s) = v_{seller}$ and $F^{-1}\left(1 - \frac{sq}{N-q}\right) = v_{buyer}$. It follows immediately that p and t are uniquely determined. ■

We can alternatively dispense with the homogenous transaction cost assumption and allow buyers and sellers to bear transaction costs $t_{buyer} \neq t_{seller}$. Though we can not jointly identify the buyer and seller specific transaction costs along with the transaction price, a pair $\tilde{p}_{buyer} = p + t_{buyer}$ and $\tilde{p}_{seller} = p - t_{seller}$ is identified. This interpretation does not affect the identification of the total transaction costs, which are $sq(\tilde{p}_{buyer} - \tilde{p}_{seller})$ either way, and would therefore not affect our later market performance calculations. Since these interpretations are observationally equivalent, we can not identify which side of the market bears the majority of the transaction costs. In the following, we therefore choose one, and maintain the homogenous transaction cost interpretation.

Our point estimates of the transaction costs and prices are given in Table 6, along with 95% confidence intervals.[11] The prices and costs are point estimates conditional on the estimated lower bound to black market trade. We show in the appendix that $\hat{t}$ is an upper bound for the transaction cost at $\hat{s}$, and that the implied $\tilde{\hat{p}}_{buyer}$ and $\tilde{\hat{p}}_{seller}$ at $\hat{s}$ are upper and lower bound estimates, respectively. The estimated demand and supply curves are plotted in Figure 9.

Our results imply that when there are material transaction costs. Trades in the black market take place between a selection of buyers with particularly high valuations and sellers with particularly low valuations. This pattern is consistent with results from the literature on the resale market of tickets where trades are observed (Bhave and Budish (2018), Leslie and Sorensen (2013)). Our estimated transaction price RMB 98 000 is higher than Li (2018)'s estimated market price about RMB 75 000 (in its Figure 3). The prices are however not directly comparable for a variety of reasons. One is that Li (2018)'s market clearing price is for a (counterfactual) market where all q licenses are offered to N households by auction, without a reservation price and with no transaction costs. Ours is a market where q households each offer its license in a market with $N - q$ buyers, and the sellers have a reservation price v_n. Our market has transaction costs, while Li's market has none. In the case of no transaction costs, our transaction price is p_{notc} = RMB 54 000.

6.1 Gains from illegal trade

We use our market performance metrics to estimate the total transaction costs and allocative gains from black market trade relative to a lottery with strictly enforced non-transferability. Our approach follows Li closely. Li's welfare calculation has two components. The first is allocative efficiency. Unlike an auction, a lottery does not allocate the licenses to the households with the highest valuations, which leads to

[11] We use the bootstrapped values of $\hat{s}$ for the estimated share of trade in Table 5 calculation for $d = 5000$.

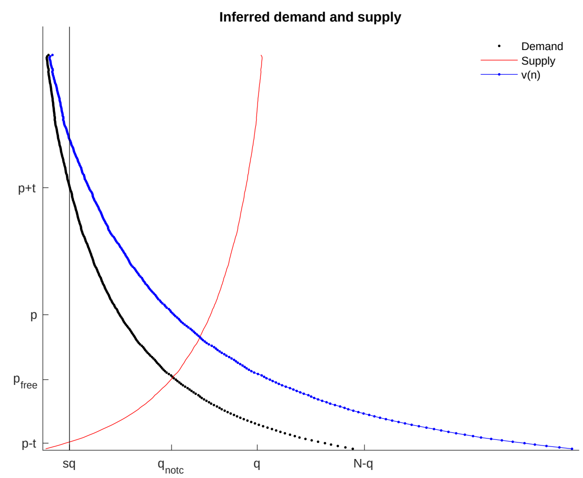

Figure 9: Estimated demand and supply. Quantity in 10 000. Price in RMB 1000.

allocative inefficiency. The second component is external pollution and congestion costs. Li shows that higher valuation households purchase larger cars with higher fuel consumption which they drive more. Therefore, a lottery increases allocative inefficiency, but reduces external costs. Li estimates the net welfare loss of the lottery relative to an auction to be about RMB 30 billion ($5 billion). The allocative inefficiency is the largest component, about RMB 33 billion, while the external cost savings are about RMB 3 billion.

Illegal trade therefore increases allocative efficiency by reallocating license plates towards households with higher valuations. Higher valuation households however also purchase larger and less fuel efficient cars that they drive more, which in turn increases congestion and pollution costs. While Li studied the welfare effects of the lottery, we study the incentives to trade, which are unaffected by externalities. We therefore ignore

Estimates	at $\hat{s} = 12\%$	95% CI lower	95% CI upper
$\hat{p}$	98.3	86.7	115.8
$\hat{t}$	93.0	79.0	113.3
total transaction costs	6.45	4.05	4.90
gains from trade	3.37	2.15	2.89

Table 6: Estimates of transaction prices and costs in RMB 1000. Gains from trade and transaction costs are in RMB billon.

the effect of the black market trade on congestion and pollution costs, but note that such trade increases externality costs.

Transaction prices are considered transfers between sellers and buyers which do not affect allocative efficiency. The allocative efficiency gain is the area between the demand and the supply curve up to $\hat{s}q$. Table 6 shows that the allocative efficiency gain is RMB 9.82 billion at $s = 12\%$, but transaction costs make up RMB 6.45 billion. A lower bound to the total gains from trade in the Beijing black market is then RMB 3.37 billion. The upper bound on the gains from trade, for the case with no transaction costs at a black market share $q_{notc} = 61\%$ and p_{notc} = RMB 54 000, is RMB 19.18 billion.

6.2 Caveats

Our derivation of the transaction costs and prices relies on Li's estimates of willingness-to-pay. These estimates are derived under the assumption of no illegal trades. Its model assumes that a prospective car buyer is of one of two types: either it needs a license to purchase a car or it does not. Since illegal trade may systematically allocate licenses to prospective buyers (households) with higher valuations, its random coefficients discrete choice estimator may not integrate over the correct distribution of household characteristics. Whether biases in the Li's demand estimates from ignoring illegal trade are economically significant is however hard to know, but if so, such biases would affect our estimates as well through $v(n)$. It is unclear how to correct for such biases, and we assume these biases away out of convenience.

We assumed that all lottery winners are drawn from the distribution of prospective car buyers. The material black market prices may however attract speculators. We define a speculator to be a household that would not take part in the lottery if non-transferability was strictly enforced. Besides the obvious economic incentive, an influx of speculators is consistent with the sharp decline in the probability of winning, which went from 10% in the first auction in January 2011, to an average of 4% in 2011, and dropping to an average of 2% in 2012.[12]

Suppose speculators have zero willingness to pay for a license plate if non-transferability is strictly enforced, i.e. a speculator will never purchase a car if she wins a license. Such speculators have two effects in our model: they crowd out car buyers on the supply side, and increase the number of car buyers on the demand side. Speculators therefore shift both the demand and the supply curve. In particular, suppose that a share z of the licenses are allocated to speculators, and suppose a speculator's reservation price is zero. Then the demand and supply is

$$D(p,t) = (N - q(1-z))(1 - F(p+t)) \tag{20}$$

$$S(p,t) = zq + (1-z)qF(p-t) \tag{21}$$

Speculators give a supply curve that is flat from zero to zq, and increasing thereafter to q. As z goes to one, the supply curve becomes vertical at q. On the demand side, the demand curve shifts out towards $v(n)$ as z goes to one as all prospective car buyers must turn to the black market for a license plate.

In Table 7, we report results assuming that all 12% illegal trades are by speculators. Since the supply curve is now flat from zero $\hat{s}q$, all speculators sell at $\tilde{p}_{seller} = 0$, and $\hat{p} = \hat{t}$.At the same time, the demand curve shifts out to $v(n)$. The outward shift in the demand curve dominates the downward shift in the supply curve: both transaction prices and costs increase, from 98.3 to 100.6 and from 93.0 to 100.6, respectively. The gains from trade go down by about 12%, largely due to further taxation of buyers.

A limit case of interest is when speculators completely crowd out prospective car buyers ($z = 1$) and there are no transaction costs. Then our model gives the same allocation as Li (2018)'s counterfactual auction market, where the supply curve is vertical

[12]The share of illegal trades may be changing over time as well, but we assume it to be constant within the period of the data.

Estimates	at $\hat{s} = 12\%$ and $z = 0.12\%$	95% CI lower	95% CI upper
$\hat{p}$	100.6	89.0	118.9
$\hat{t}$	100.6	89.0	118.9
total transaction costs	7.00	4.36	5.25
gains from trade	2.98	1.84	2.48

Table 7: Estimates of transaction prices and costs at $z = 0.12$. Prices and costs are in RMB 1000 . Gains from trade and total transaction costs in RMB billion

at q and demand is $v(n)$. It follows that Li (2018)'s analysis applies.[13]

7 Discussion

It is puzzling that the Beijing government seems to leniently enforce the non-transferability of the license plates. One reason may be that the scale of trade in the license plates implies prohibitive enforcement costs. The lax enforcement may also strike a balance between allocative inefficiency and equity concerns in a society where only 8% of the population approves of market based allocations like auctions (Li (2018)). Some illegal trade may not take place in a market place, but is plausibly between family members, e.g. a daughter wins a license and lets her parents register a car in her name. Such trades may be less politically expedient to enforce.

Governments sometimes welcome black markets. Chinn (1977) notes that during the rationing of rice in Japan post World War II, the Japanese government not only allowed a sizable black market, estimated to be about half of the total market, but also collected data on prices and quantities traded illegally. The government price controls were set to dampen industry wages in a time of rapid industrialization, while the black market served as a slackness control to ameliorate the most severe allocative inefficiencies. Though it is hard to argue that a leniently enforced lottery is an optimal mechanism, the lottery may serve its purpose better than an auction in a city like Beijing where the public support for market based allocations is found to be weak.

Perhaps the enforcement is not as lenient after all. The upper bound on the transaction costs, which total about twice the transaction price, suggest severe market frictions.

[13]Except for the congestion and pollution externalities, which we have ignored.

The high willingness to pay for license plates in the market by itself creates strong incentives for illegal trade, so the size of the implied transaction costs suggests that enforcement in fact precludes trades of fairly high value, but still admits a sizable share of high value trades. If interest is in a well functioning black market, the upper bound on the transaction costs suggest relaxing enforcement. If interest is in enforcing the non-transferability, the upper bound suggest that enforcement is possibly quite effective in precluding illegal trade of high private value. The results are however also consistent with no transaction costs, i.e. completely ineffective enforcement, and black market trade of 61% of the quota.

8 Summary

Exploiting the lottery allocation of license plates along with minimal assumptions on households' car preferences, we applied an optimal transport model to estimate a lower bound of the share of license plates that were illegally traded in the black market. Combining the estimated black market share with Li (2018)'s welfare estimates in a black market equilibrium model, we derived transaction prices and costs. These imply a conservative estimate of the gains from trade in the black market to be RMB 3.37 billion, or about $0.56 billion, yet up to RMB 66% are lost to transaction costs.

In our application, optimal transport provides the natural summary statistic for the analysis at hand, and may be used as easily as one would use standard regression analysis. The estimators follow as solutions to well-behaved optimization problems and are computationally light even for moderately big data sets. We believe optimal transport can be useful as a complement to standard event study approaches, in particular where credible estimates are important for policy.

References

Bhave, A. and E. Budish (2018). Primary-market auctions for event tickets: Eliminating the rents of 'bob the broker'? Working Paper w23770, NBER. 26, 29

Carlier, G., V. Chernozhukov, and A. Galichon (2016, 06). Vector quantile regression: An optimal transport approach. *The Annals of Statistics 44*(3), 1165–1192. 10

Chiappori, P.-A. and B. Salani (2016). The econometrics of matching models. *Journal of Economic Literature 54*(3), 832–861. 3

Chinn, D. L. (1977). Staple food control and industrial development in postwar japan, 19501957: The role of the black market. *Journal of Development Economics 4*(2), 173 – 190. 33

Chiong, K., A. Galichon, and M. Shum (2016, 3). Duality in dynamic discrete-choice models. *Quantitative Economics 7*(1), 83–115. 10

Davidson, C., L. Martin, and J. D. Wilson (2007). Efficient black markets? *Journal of Public Economics 91*(7), 1575 – 1590. 2

Dye, R. A. and R. Antle (1986). Cost-minimizing welfare programs. *Journal of Public Economics 30*(2), 259 – 265. 2

Feige, E. (2016, 06). Professor schneider's shadow economy:what do we really know? a rejoinder. *Journal of Tax Administration 2*, Forthcoming. 2

Galichon, A. (2016). *Optimal transport methods in Economics*. Princeton University Press. 3

Galichon, A., S. D. Kominers, and S. Weber (2018). Costly concessions: An empirical framework for matching with imperfectly transferable utility. *Journal of Political Economy 0*(ja), null. 10

Gallagher, A. W. A., K. A. Evans-Reeves, J. L. Hatchard, and A. B. Gilmore (2019). Tobacco industry data on illicit tobacco trade: a systematic review of existing assessments. *Tobacco Control 28*(3), 334–345.

Kantorovitch, L. (1958). On the translocation of masses. *Management Science 5*(1), 1–4. 10

Leslie, P. and A. Sorensen (2013, 10). Resale and Rent-Seeking: An Application to Ticket Markets. *The Review of Economic Studies 81*(1), 266–300. 29

Li, S. (2018, 10). Better Lucky Than Rich? Welfare Analysis of Automobile Licence Allocations in Beijing and Shanghai. *The Review of Economic Studies 85*(4), 2389–2428. 3, 4, 5, 8, 11, 12, 26, 27, 29, 30, 31, 32, 33, 34, 37

Peyré, G. (2018). Mathematical foundations of data sciences. *https://mathematical-tours.github.io*, 267–297. 21, 22

Schneider, F., A. Buehn, and C. E. Montenegro (2010). New estimates for the shadow economies all over the world. *International Economic Journal 24*(4), 443–461. 2

Sommerfeld, M. and A. Munk (2018). Inference for empirical wasserstein distances on finite spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(1), 219–238. 24

Stahl, D. O. and M. Alexeev (1985). The influence of black markets on a queue-rationed centrally planned economy. *Journal of Economic Theory 35*(2), 234 – 250. 2

Tan, J., J. Xiao, and X. Zhou (2019). Market equilibrium and welfare effects of a fuel tax in china: The impact of consumers' response through driving patterns. *Journal of Environmental Economics and Management 93*, 20 – 43. 4, 8, 12

Thomas, J. (1999). Quantifying the black economy: 'measurement without theory' yet again? *The Economic Journal 109*(456), F381–F389. 2

Tobin, J. (1952). A survey of the theory of rationing. *Econometrica 20*(4), 521–553. 11

Xiao, J., X. Zhou, and W.-M. Hu (2017). Welfare analysis of the vehicle quota system in China. *International Economic Review 58*(2), 617–650. 3, 4, 8, 12

A Comparative statics in transaction costs and prices

Taking derivatives of the equilibrium conditions (19), we get

$$\begin{aligned} 2\frac{\partial p(s)}{\partial s} &= \frac{1}{f(v_{seller})} - \frac{q}{N-q}\frac{1}{f(v_{buyer})} \\ 2\frac{\partial t(s)}{\partial s} &= -\frac{q}{N-q}\frac{1}{f(v_{buyer})} - \frac{1}{f(v_{seller})} \end{aligned}$$

It is immediately clear that the inferred transaction costs decrease with the estimated size of the black market. The same is not necessarily true for the transaction price, which may increase or decrease, depending on the shape of f. Both $\tilde{p}_{buyer} = p + t$ and $\tilde{p}_{seller} = p - t$ are monotonic in s. These statics show that our estimate $\hat{t}$ and $\tilde{\hat{p}}_{buyer} = \hat{p} + \hat{t}$ are upper bound estimates, and that $\tilde{\hat{p}}_{seller} = \hat{p} - \hat{t}$ is a lower bound estimate. Since both p and s are equilibrium objects, while the transaction costs are exogenous by assumption, these comparative statics are not meaningful in the economic analysis.

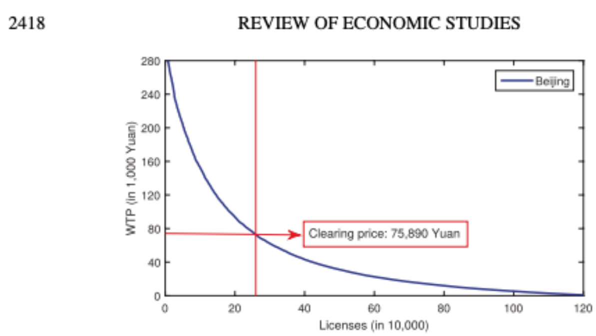

Figure 10: Li's original Figure 4

B A general difference-in-difference optimal transport estimator (TBD)