CENTRE *for* ECONOMIC
PERFORMANCE

CEP Discussion Paper No 1646

August 2019

Does E-Commerce Reduce Traffic Congestion?
Evidence from Alibaba Single Day Shopping Event

Cong Peng

THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE ■

Abstract
Traditional retail involves traffic both from warehouses to stores and from consumers to stores. E-commerce cuts intermediate traffic by delivering goods directly from the warehouses to the consumers. Although plenty of evidence has shown that vans that are servicing e-commerce are a growing contributor to traffic and congestion, consumers are also making fewer shopping trips using vehicles. This poses the question of whether e-commerce reduces traffic congestion. The paper exploits the exogenous shock of an influential online shopping retail discount event in China (similar to Cyber Monday), to investigate how the rapid growth of e-commerce affects urban traffic congestion. Portraying e-commerce as trade across cities, I specified a CES demand system with heterogeneous consumers to model consumption, vehicle demand and traffic congestion. I tracked hourly traffic congestion data in 94 Chinese cities in one week before and two weeks after the event. In the week after the event, intra-city traffic congestion dropped by 1.7% during peaks and 1% during non-peak hours. Using Baidu Index (similar to Google Trends) as a proxy for online shopping, I found online shopping increasing by about 1.6 times during the event. Based on the model, I find evidence for a 10% increase in online shopping causing a 1.4% reduction in traffic congestion, with the effect most salient from 9am to 11am and from 7pm to midnight. A welfare analysis conducted for Beijing suggests that the congestion relief effect has a monetary value of around 239 million dollars a year. The finding suggests that online shopping is more traffic-efficient than offline shopping, along with sizable knock-on welfare gains.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

1   Introduction

As the digital economy takes shape, e-commerce continues to grow around the world. The past decade has seen an explosive growth of the Chinese e-commerce market, with the Gross Merchant Value increased by more than ten times[1]. Today, more than 40% of the world's e-commerce transactions take place in China, a dramatic increase from only 1% about a decade ago. In the year 2016, there were about 460 million online consumers and the share of online consumption is about 12.6% of total consumption. Traditional retail involves traffic both from warehouses to stores and from consumers to stores. E-commerce cuts intermediate traffic by delivering goods directly from the warehouses to the consumers and improve the efficiency of logistics of goods. This poses the question of whether e-commerce is more traffic efficient than traditional retail. This paper investigates the possible reduction of traffic channeled through the rapid expansion of e-commerce and its new logistics.

Although plenty of evidence has shown that vans that are servicing the e-commerce are a growing contributor to traffic and congestion, consumers are also found making less shopping trips using vehicles (Braithwaite, 2017). The trade-off between the two effects is crucial to assess the overall effect of e-commerce on traffic. Punakivi (2003) simulated the replacement of traditional retailing by electronic retailing and found that this potentially leads to 54-95% reduction in traffic depending on delivery methods. In a similar vein, Cairns (2005) estimates that a direct substitution of car trips by van trips could reduce vehicle-km by at least 70%. A recent comprehensive report by Braithwaite (2017) gathered suggestive evidence indicating that online shopping is likely to reduce overall shopping traffic in probably modest scale in the real world. Measuring congestion reduction is difficult because of latent travel demand suppressed by traffic congestion itself. When traffic congestion was reduced, individuals who chose not to travel may decide to travel after observing less traffic. In fact, the reduction of traffic congestion might never be observed in the long term due to the fundamental law of road traffic congestion (Duranton and Turner, 2011).

In this paper, I provide the first available estimates of the effect of e-commerce on traffic and congestion. The analysis utilizes temporary price shocks caused by a nationwide online shopping event in 2016 as the foundation of my identification strategy. The Singles' Day shopping event on the day of 11 November each year is the largest online sales event in China, equivalently as popular as the shopping event of Cyber Monday in the United States[2]. In the year 2016, the largest Chinese

---

[1]According to the Chinese E-commerce Association. `http://www.100ec.cn/zt/upload_data/2018dzswfzbg.pdf`

[2]The concept of Singles' Day was initiated by young college students to buy presents to celebrate single-hood and resist social pressure to get married. 11 November was chosen as the date because each of its lone digits (11.11) represents a "bare stick" as the symbol of uncoupled individuals. The day was then promoted by Alibaba as a cultural

online shopping platform, Alibaba, reached a new sales record at 17.6 billion US dollars during a 24-hours promotion period. To put this figure into context, the annual online consumption is 587 billion US dollars. The average price in the online channel on the event day is about 80% of the average price in the month before the event[3].

A reduced price in the online channel during the sale event encourages consumers to switch from offline channel to online channel. I measure the change in traffic congestion in each hour one week before and two weeks after the event. Traffic congestion is measured by an index that is the ratio of actual passing time and free-flow passing time of vehicles in a road segment. The road-level information is first collected by Global Positioning System (GPS) trackers based on millions of users of a navigation service company, then aggregated to city-hour level, and finally, released for public use. Change in online shopping is measured by Baidu Index, which is similar to Google Trends but can track search Internet Protocol (IP) addresses to cities. In the week after the event, intracity traffic congestion dropped by 1.7% during peak hours and 1% during off-peak hours, while online shopping increased by about 1.6 times. Cities with a higher increase in online shopping experienced a greater reduction in traffic congestion.

How does online shopping affect traffic congestion? To answer this question, I derive the relationship between online consumption quantity, vehicle demand, and a traffic congestion index. Based on the speed-density relationship proposed in Adler et al. (2017), the logarithm of the traffic congestion index can be expressed as the ratio of vehicle density to the free-flow vehicle density. The change in the logarithm of the traffic congestion index is decided by the change in traffic density, which in turn is decided by the change in vehicle demand. If the vehicle demand per unit goods of online shopping is only a fraction of that of offline shopping, which I call a "vehicle-saving ratio", then the total vehicle demand for shopping changes with the online-offline substitution of consumption induced by the price shock. With a one unit increase in online consumption, the net reduction in traffic is the difference between the following two parts: the reduction in offline shopping vehicle subject to the online-offline substitution of consumption and the share of shopping made through private vehicles, and the increase in vehicle demand arising from online shopping. I call this net reduction a "traffic-saving factor". In addition, I measure the online-offline substitution using the ratio of the increase in online consumption over the reduction in offline consumption under a price shock, which I call a "online-offline substitution ratio". With simple accounting of traffic, the model further reveals that the elasticity of traffic congestion index to online consumption quantity is the traffic-saving factor adjusted by the current congestion level and a function of three shares:

event of online shopping, bearing similarity with Cyber Monday in the United States.

    [3]The online shopping platform did not publish relevant price change information. An independent credit rating company, CCX Credit Technology, monitored the price of a sample of online products surrounding the event and published the findings online (see `http://www.01caijing.com/article/12269.htm`)

the share of online shopping, the share of shopping made through private vehicles, and the share of vehicles used due to shopping. Particularly, the condition for online shopping to reduce traffic is the traffic-saving factor being negative. To be specific, the condition requires that the vehicle-saving ratio is sufficiently low, or the amount of offline shopping that consumers are willing to substitute with online shopping is sufficiently large. The first part can be estimated from an operations management perspective using data relatively accessible. For example, the vehicle-saving ratio can be estimated using data on the number of online goods delivered in a van in an hour, and the number of offline goods purchased in a private vehicle[4]. However, the second part requires data that are more difficult to obtain[5]. For this reason, I developed a demand model to predict the online-offline substitution ratio.

The model is characterized by two key assumptions. First, the utility that a consumer obtains from a product depends on the matching quality between the consumer and the product. The matching quality is assumed to be a random variable that follows the Fréchet distribution. The quantity of consumption adjusted by matching quality is aggregated across products by a CES functional form to derive a consumer's utility. Second, I assume a mechanism of how consumers choose shopping channels for a product. The matching quality is assumed to be the maximum of two underlying channel-specific draws of matching quality. Specifically, given a product, a consumer draws an online matching quality from an online Fréchet distribution, and then draw an offline matching quality from an offline Fréchet distribution. The consumer then chooses the maximum of the two draws, and thus self-selects into a type of either online consumer or offline consumer based on the channel that yields the maximum draw[6]. I assume that the two channel-specific Fréchet distributions have the same shape (which decides variability) but different scales (a higher scale means that a high-value draw is more likely). The relative value of the channel-specific scale parameters, along with the shape parameter, determine the probability under which consumers choose each channel. Intuitively, the channel with a larger scale parameter attracts a higher proportion of consumers.

As the maximum of two Fréchet random variables also follows a Fréchet distribution[7], I can derive the overall demand (sum of demand from the two channels) for a specific product. The expectation of the overall demand for the product is used by monopolist firms to set the equilibrium price, which is the marginal cost of the product multiplied by a constant mark up – a well-known

---

[4]I provide a crude estimate in Section 2.3 based on this simple logic. The vehicle-saving ratio is about 0.067.

[5]The data required to measure the online-offline substitution ratio empirically are unavailable. It requires weekly data on the online and offline consumption surrounding the event day. Even if the online shopping platform would like to share the online consumption surrounding the event, obtaining offline shopping consumption on a weekly basis remains challenging. Potentially, I could conduct a consumer survey, but the large data requirements are outside the scope of this study.

[6]The same consumer can be an online consumer for one product and an offline consumer for another product.

[7]The parameters of the former can be derived from the parameters of the latter.

result under the assumptions of CES utility and monopolistic firms. Given the fixed price of a specific product, consumers divide into online or offline consumers according to the probability under which consumers choose channels. Conditional on being online (or offline) consumers, the quantity consumed within these online (offline) consumers turns out to follow the same distribution as the overall quantity consumed by all consumers. In other words, the demand distribution in each channel is independent of the channel. This is similar to a key finding in Eaton and Kortum (2002)[8]. I find that this independence property has two important implications[9]:

  (i) The expected values of the quantity consumed in each channel are equal.

 (ii) The expected channel-specific quantity consumed equals the share of consumers that choose that channel multiplied by the expected overall demand from both channels.

Importantly, these properties hold when prices change. As a result, the change in channel-specific demand due to price shock can be approximated by a derivative formula. This allows for calculating the ratio of the increase in online shopping quantity over the reduction in offline shopping quantity, which gives the online-offline substitution ratio. Further, the ratio can be expressed as a concise function of the elasticity of online shopping quantity to the relative price of online to offline channel, and the elasticity of substitution between varieties. The former can be estimated from the data, and the latter is a well-studied parameter in the trade literature.

Armed with the formula of the online-offline substitution ratio, I conduct a quantitative analysis to explore whether the condition for online shopping to reduce traffic congestion holds and the quantitative importance of the elasticity of traffic congestion to online consumption quantity, with reasonable guesses on the parameters and sample statistics of variables in the model. Particularly, the online-offline substitution ratio is estimated to be around -1.9, which indicates that the reduction in offline consumption due to one unit increase in online shopping is about a half unit. Several well-educated guesses on the parameters in the quantitative analysis show that the condition is likely to hold. However, the elasticity of traffic congestion to online consumption quantity appears to have a wide range of values due to heterogeneity. For example, the estimates vary from -0.06 to -0.25 for different cities. Therefore, in order to identify an average magnitude of the elasticity, I turn to empirical estimates.

In the empirical section, Ordinary Least Square (OLS) regression models estimate that a 10% increase in online shopping reduces traffic congestion by about 0.13%, an elasticity of -0.013. The

---

[8]One of the key findings in Eaton and Kortum (2002) is that the price distribution of the varieties that any given origin actually sends to any given destination is independent of its origin regions.

[9]Note these properties rely heavily on the assumptions of the Fréchet distribution.

effect mostly comes from peak hours, that is from 9 am to 11 am and from 7 pm to midnight. To address potential endogeneity problems arising from omitted variable bias, I instrument the change of online shopping in the event with the reduction in postage fee. Postage fee was waived on the day of the shopping event and introduces exogenous incentives for consumers to switch to online shopping conditional on market access. Using the waived online delivery postage fee as the instrumental variable (IV), the IV estimation reveals much stronger effects than the OLS estimates. A 10% increase in online shopping reduces traffic congestion by about 1.4%, an elasticity of -0.14, which is consistent with the range from the quantitative analysis. The weights in the IV estimates appear to be assigned towards cities that experience a higher increase in online shopping. Using the IV estimates to conduct a welfare analysis for Beijing, the congestion relief effect of 10% increase in online shopping can be converted to monetary terms of 239 million US dollars a year for peak hour commuters, which is equivalent to about a third of the average effect of providing access to an additional subway line, according to the welfare gains from the congestion relief effect of subways estimated by Gu et al. (2019).

This paper contributes to a growing literature in the spatial economic impact of the digital economy, as reviewed in Goldfarb and Tucker (2019). As a digital purchasing technology, e-commerce reduces transportation cost. On the consumption side, consumers choose online shopping to reduce travel cost despite certain disutility in online shopping (Forman et al., 2009). E-commerce also contributes to overcoming the logistical barrier in rural areas and leads to sizable gains in real household income. Consumers in villages benefit from greater product variety and lower prices driven by a significant reduction in travel costs (Goolsbee and Klenow, 2018; Couture et al., 2018). Consequently, e-commerce reduces spatial inequality of consumption between large cities and small cities, and increases access to varieties (Dolfen et al., 2019; Fan et al., 2018). On the supply side, online shopping causes structural changes in offline shopping economies. It shifts the market from high-cost producers to low-cost producers (Goldmanis et al., 2010). In the retail industry, offline retailers that directly compete with online retailers are negatively impacted, while indirect competitors that can adapt the change in e-commerce revolution and take advantage of the online-offline complementarities can be winners in the competition (Relihan, 2017). Dolfen et al. (2019) concludes that consumer gains are about 1.1% of all consumption using credit card data in the US, which is tantamount to 1,150 US Dollars per household in the year of 2017. This paper advances our understanding of the impact of e-commerce on urban traffic congestion.

This paper is also linked to a long literature in traffic congestion relief policies. Many policy options have been extensively studied in the past, for example, congestion charge (Yang et al., 2018a; Tikoudis et al., 2015) and quantity-based restriction (Yang et al., 2014) on the demand side; transport infrastructure expansion and subsidies (Parry and Small, 2009; Anderson, 2014; Yang et

al., 2018b; Gu et al., 2019) on the supply side. These measures are either politically controversial or expensive (Adler et al., 2017). The promotion of e-commerce is considered a "soft" policy that seeks to encourage people to reduce their car usage through enhancing the awareness and attractiveness of alternative options (Cairns et al., 2004).

The model in this paper draws insights from three strands of models. The first strand is a long list of trade models with constant elasticity demand (Armington, 1969; Krugman, 1980; Eaton and Kortum, 2002; Helpman et al., 2004; Antràs et al., 2017). The elements in these models are helpful in understanding trade flows across locations. Particularly, Fan et al. (2018) and Dolfen et al. (2019) have applied this type of trade model to analyze e-commerce. The second strand of models is in the literature of marketing economics and industrial organization. One key insight is that firms can employ random sales to compete over consumers with lower willingness to pay and reach higher profit, relative to only serving a fraction of consumers with higher willingness to pay (Varian, 1980; Seim and Sinkinson, 2016). However, their study does not allow for consumers with continuously distributed taste. Drawing insights from random coefficients demand models such as Coşar et al. (2018) and Berry et al. (1995), I introduced heterogeneous consumers into the demand side but specify their tastes for channels following Fréchet distribution [10]. Different to their study, I focus on the online retail event and analyze the substitution between online and offline products when the relative prices change during a sale. The third strand includes models developed by transportation engineers that study the relationships between speed, density and flow. Particularly, I adopt the speed-density relationship proposed as in Adler et al. (2017), and derive an exact functional form that links the change of online consumption, vehicle demand and traffic congestion index.

This paper is structured as follows. Section 2 sets out a model to connect the change in online shopping in the event with the change in traffic congestion. Section 3 outlines data and shows a descriptive analysis. Section 4 delves into the econometric framework. Section 5 presents initial evidence on the link between the increase in online shopping and the reduction in traffic congestion. Section 6 presents the causal estimates of the effect of online shopping on traffic congestion. Section 7 analyzes the welfare impact of congestion relief effect of e-commerce. Section 8 discusses the long-term effect. Section 9 concludes.

2 Theoretical Framework

This section provides a theoretical framework to link the price change in online shopping and traffic congestion. The first part of the section lays out a demand model to predict the online-offline substitution. The analysis of the quantity demanded in both channels draws from Eaton and

---

[10]Redding (2016) specified worker's taste for amenity to follow Fréchet distribution.

Kortum (2002)'s approach in analyzing the price distribution across origin countries. Although I use a multiple regions trade model framework, channel choices and online-offline substitution are the main focus. The key insight of the model remains even if reducing the multiple regions in the model to a single city. The second part of the section shows the derivation of the exact functional form for the relationship between traffic congestion index and online shopping quantity, and the condition for online shopping to reduce traffic congestion. The last part of the section explores the quantitative importance of the congestion relief effect of online shopping based on the model. In contrast to the introduction section, I start with the micro-foundation of the model and then move on to the aggregate consumption by channels in cities and its relationship with traffic congestion, which is a macro-phenomenon.

2.1 Demand Model and Online-Offline Substitution of Consumption

I develop a demand model to quantify the online-offline substitution of consumption (quantity). On the demand side, I allow for heterogeneous consumers. For a given variety[11], Fréchet distributed consumer-variety matching quality separates consumers into two types: online and offline consumers. On the supply side, I assume that each firm produces a unique variety and acts as a monopolist. It prices the variety uniformly in the two market channels, based on expected demand. In order to analyze a series of changes in the economy during the sale event, I introduce exogenous price shocks in both channels so that firms can price discriminate temporarily[12] across the two channels. When online prices decrease, online consumers will consume more, and some offline consumers will switch their types to online consumers. As a result, the online shopping event alters the share of online shopping and offline shopping and the overall amount of online shopping at the aggregate level. The ratio of the increase in online shopping quantity over the reduction in offline shopping quantity gives the online-offline substitution of consumption.

2.1.1 The Set-Up

There are I cities, with each city indexed by i or j depending on whether the region in question is the origin, i, or the destination, j, of a trade. There is a continuum of consumers indexed by μ in city j, with the set of consumers U_j and the mass of consumers L_i. Each city is endowed with L_i units of workers (the same as the the mass of consumers) where each worker supplies one unit of labor inelastically and receives wage w_i. Suppose that labor is the only factor of production. Consumers buy (or firms sell) from online and offline channels, indexed by $m \in \{o, f\}$. Each variety is indexed by ω. Suppose that every firm in the world produces a distinct variety. The concept of a city is thus

---

[11]I use the word variety and product interchangeably in the paper.

[12]There was no parallel offline sales during the event in 2016. Competing offline sales started in 2017.

a cluster of firms with the same productivity but distinct products, and consumers with the same income but different tastes for online shopping.

2.1.2 Consumers

Assume individuals obtain utility $z(\mu, \omega)q_j(\mu, \omega)$ from consuming $q_j(\mu, \omega)$ units of variety ω, where $z(\mu, \omega)$ is the matching quality for each pair of consumer μ and variety ω. Consumers maximize a CES objective:

$$u_j(\mu) = \left(\int_{\Omega_j} (z(\mu, \omega)q_j(\mu, \omega))^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{1}$$

I assume that the unobserved matching quality is a random component and follows the Fréchet (Type-II Extreme Value) distribution. This idea is fundamentally similar to Redding and Weinstein (2016) where their CES preference parameters are assumed to vary across both product type and consumer type[13][14][15].

I assume $z(\mu, \omega)$ is generated following the process below. Given variety ω, each consumer receives two draws $z_o(\mu, \omega)$ and $z_f(\mu, \omega)$ from two Fréchet distributions $F(\theta, s_o(\omega))$ and $F(\theta, s_f(\omega))$ for online and offline channels, respectively. The Fréchet distribution for channel m is:

$$F_m(z) \equiv Pr\{z_m(\mu, \omega) \le z\} = exp\left\{-\left(\frac{z}{s_m(\omega)}\right)^{-\theta}\right\} \tag{2}$$

where θ is the shape parameter. It governs the amount of variation within the distribution. $s_m(\omega)$ captures general preference for a specific channel and is assumed to be the same for all consumers but different across varieties (thus not index by μ). A bigger $s_m(\omega)$ implies a higher draw of matching quality $z_m(\mu, \omega)$ for any variety ω in channel m is more likely.

$$q_{mj}(\mu, \omega) = Y_j P_j^{\sigma-1} \left(\frac{p_j(\omega)}{k_m}\right)^{-\sigma} z_m(\mu, \omega)^{\sigma-1} \tag{3}$$

where $p_j(\omega)$ is the price of variety ω in city j. It is the normal price listed by the retailers without considering any sales events. As shown later in this model, the price is marginal cost adjusted by

---

[13]Their preference parameters are not random variables, but the authors introduce a Fréchet distributed shock as the multiplier to preference parameters, so the composite is equivalent to z here.

[14]As summarized in Redding and Weinstein (2016), heterogeneous random utility models have been studied in demand system estimation literature such as Coşar et al. (2018), Berry et al. (1995), McFadden (1973), etc.

[15]McFadden (1973) argues that consumers select the products that maximize their utility from a set of alternatives. Kortum (1997) shows a model where research leads to draws from a Pareto distribution, causing the technological frontier to be distributed Fréchet. As a conjecture, if consumers search for the best products in suiting their preferences, the search process may give rise to the Fréchet distribution of the matching quality.

a mark-up. k_m represents a channel-specific price shock[16], which captures any temporary changes of prices due to sales on the supply side. P_j is price index in city j[17],

$$P_j = (\int_{\Omega_j} (\frac{p_j(\omega)}{z_{max}(\mu, \omega)})^{1-\sigma})^{\frac{1}{1-\sigma}} \tag{4}$$

In a nutshell, consumers choose the shopping channel which gives higher $q_{mj}(\mu, \omega)$. This means that consumers compare $k_o^\sigma z_o(\mu, \omega)^{\sigma-1}$ with $k_f^\sigma z_f(\mu, \omega)^{\sigma-1}$ and pick the channel that yields higher value. The quantity consumed is the max of $q_{oj}(\mu, \omega)$ and $q_{fj}(\mu, \omega)$. The assumed process above provides the mechanism to decide the type of the consumer (online versus offline) and the probability of making such choices. Consumer μ in city j maximizes utility under the constraint of income Y_j. Here, consumers in the same city are assumed to have the same income, for model simplicity. The demand function of an individual consumer μ is,

$$q_j(\mu, \omega) = max\{q_{oj}(\mu, \omega), q_{fj}(\mu, \omega)\} \tag{5}$$

Given the distribution of channel specific matching quality, $z_m(\mu, \omega)$, the distribution of $q_j(\mu)$ in consumers for variety ω in channel m is

$$G_{mj}(q) \equiv Pr\{q(z_m(\mu, \omega)) \leq q)\}$$
$$= exp\{-(\frac{q}{p_j(\omega)^{-\sigma} Y_j P_j^{\sigma-1} (k_m^{\frac{\sigma}{\sigma-1}} s_m(\omega))^{\sigma-1}})^{-\frac{\theta}{\sigma-1}}\} \tag{6}$$

Consumers choose the channel that *potentially* gives the higher quantity, so the distribution of $q_j(\mu)$ that consumers *actually* buys from either channel for variety ω in city j is:

$$G_j(q) \equiv Pr\{max\{q_o(\mu, \omega), q_f(\mu, \omega)\} \leq q\} \tag{7}$$
$$= exp\{-(\frac{q}{p_j(\omega)^{-\sigma} Y_j P_j^{\sigma-1} ((k_o^{\frac{\sigma}{\sigma-1}} s_o(\omega))^\theta + (k_f^{\frac{\sigma}{\sigma-1}} s_f(\omega))^\theta)^{\frac{\sigma-1}{\theta}}})^{-\frac{\theta}{\sigma-1}}\}$$

The calculation of the aggregated demand for variety ω in city j requires integrating individual demand function on the probability distribution of $z_m(\mu, \omega)$, which is equivalent to calculating the expectation of $q_j(\mu, \omega)$. Using the mean function of the Fréchet distribution, the overall demand

---

[16]k_m is defined in the way that k_o increases from 1 to about 1.25, when there is 20% online discount.

[17]I omit the k_m from the price index for three reasons: First, k_m is always one when there is not a sales event. Second, I assume that the price shock will not be large enough to affect the overall price index. Third, consumers perception of price index is unlikely to adjust in a week. In other words, consumers do not feel becoming relative richer because of the sales event.

for variety ω in city j is:

$$Q_j(\omega) = \int_{\mu \in U_j} q_j(\mu, \omega) d\mu \tag{8}$$

$$= L_j p_j(\omega)^{-\sigma} Y_j P_j^{\sigma-1} \Gamma(1 - \frac{\sigma-1}{\theta})((k_o^{\frac{\sigma}{\sigma-1}} s_o(\omega))^\theta + (k_f^{\frac{\sigma}{\sigma-1}} s_f(\omega))^\theta)^{\frac{\sigma-1}{\theta}}$$

where Γ is the Gamma function. Note that the shape parameter that governs the dispersion of matching quality θ is the same for both channel-specific distribution of $z_m(\mu, \omega)$, and the distribution of $z(\mu, \omega)$, which is the matching quality in the chosen channel. This again relies on the properties of the Fréchet distribution. Mathematically, the expectation exists only when $\theta > \sigma - 1$[18]. This means that the dispersion of $z(\mu, \omega)$ should be large enough so that the dispersion of matching quality is larger than the elasticity of substitution between varieties[19].

2.1.3 Firm Decisions

Without considering any sales event, firms price two channels uniformly to avoid arbitrage[20]. A firm is characterized by its productivity parameter of ϕ_i. Firms are assumed to have dual channels in the sense that all firms in city i can sell through both online channel and offline channel to consumers in city j[21]. Therefore, firms optimize price based on the aggregate demand in each city, assuming that firms have such information.

$$W_{ij}(\omega) = (p_{ij}(\omega) - \frac{w_i \tau_{ij}}{\phi_i}) Q_{ij}(\omega) \tag{9}$$

where $W_{ij}(\omega)$ is the profit, w_i is the wage in city i, ϕ_i is the productivity in city i, and $Q_{ij}(\omega)$ is the expected demand for ω in city j. τ_{ij} is the iceberg transport cost to ship the goods from city i to j. For simplicity, intercity transportation costs are assumed equal across the two channels. This is plausible as moving goods across cities in both channels use the same technology, which is mostly rail freight or highway freight transport. Another reason is that I do not have data on the freight cost. Given the assumption that firms price based on the expected aggregate demand in a city, so

---

[18]The expectation of a random variable with Fréchet distribution exists only if the scale parameter of the distribution is larger than 1.

[19]Given that the dispersion of matching quality measures the variability of alternative choices (in my context, shopping channels), I interpret this condition as that the variability of alternative choices for a variety is larger than the substitutability between the chosen varieties.

[20]Cavallo (2017) shows that online and offline prices are identical about 72% of the time of study period from December 2014 to March 2016.

[21]I only focus on dual channel firms because the event only allows small shops (with both online and offline distribution channels) to participate. In a similar vein to Fan et al. (2018) and Helpman et al. (2004), it is possible to derive the fraction of online-only firm and dual-channel firms based on the trade-off between additional revenue from satisfying a greater variety of consumers in taste for channels and the additional cost in setting up physical stores.

that the uncertainty of individual consumption induced by $z(\mu, \omega)$ does not affect the well-known, constant mark-up under CES utility function and monopolistic competition[22]. The optimal price is,

$$p_{ij}(\omega) = \frac{\sigma}{\sigma - 1} \frac{w_i \tau_{ij}}{\phi_i} \tag{10}$$

Channel-specific price shock k_m is not included here because the long-term price does not include discounts. Discount in the sales event is treated as an external shock in the model[23]. The model does not include the dynamics that consumers may anticipate the event and shift their budget for three reasons. First, given that this is a yearly event, and the exact rules and the products on sale change every year, there are many unanticipated elements. Second, the goal of the model is to analyze online-offline substitution corresponding to price change. Including those dynamics may overly complicate the model without adding much insight. Third, I address the anticipation effects separately in the empirical section.

2.1.4 Equilibrium Quantity and Share of Consumers by Channels

The expected equilibrium consumption of good from city i in city j is,

$$\begin{aligned}
Q_{ij}(\omega) &= L_j E(q(\mu, \omega)) \\
&= L_j (\frac{\sigma}{\sigma - 1} \frac{w_i \tau_{ij}}{\phi_i})^{-\sigma} Y_j P_j^{\sigma-1} \Gamma(1 - \frac{\sigma-1}{\theta})(k^{\frac{\sigma}{\sigma-1}} s_f(\omega))^{\sigma-1}(1 + (k^{\frac{\sigma}{\sigma-1}} s(\omega))^{\theta})^{\frac{\sigma-1}{\theta}} \equiv C_{ij} A
\end{aligned} \tag{11}$$

where $k \equiv \frac{k_o}{k_f}$, $s(\omega) \equiv \frac{s_o(\omega)}{s_f(\omega)}$. Here, I normalize the price shocks and average preference for both channels by the values in the offline channel to reduce the number of unknown variables. For simplicity, let $A \equiv (1 + (k^{\frac{\sigma}{\sigma-1}} s(\omega))^{\theta})^{\frac{\sigma-1}{\theta}}$, and C_{ij} denotes a collection of items not related to k and $s(\omega)$.

Given the probability of $z_m(\mu, \omega)$, we can calculate the probability that a consumer buys through channel m. Given the law of large numbers, the probability gives the fraction of consumers that choose channel m,

$$\pi_{oj}(\omega) = Pr\{k_o^{\frac{\sigma}{\sigma-1}} z_o(\mu, \omega) \geq k_f^{\frac{\sigma}{\sigma-1}} z_f(\mu, \omega)\} = \frac{(k^{\frac{\sigma}{\sigma-1}} s(\omega))^{\theta}}{1 + (k^{\frac{\sigma}{\sigma-1}} s(\omega))^{\theta}} \tag{12}$$

---

[22]Gabaix et al. (2016) presents an interesting discussion on how noise in prices may affect mark-up.

[23]Discount can be modeled endogenously as a part of the pricing strategy as in Seim and Sinkinson (2016). However, since the goal of my model is to provide a framework to evaluate the ratio of the increase of sales in the online channel to the decrease of sales in the offline channel, endogenous discounts are beyond the scope of this study.

$$\pi_{fj}(\omega) = 1 - \pi_{oj}(\omega) = \frac{1}{1 + (k^{\frac{\sigma}{\sigma-1}} s(\omega))^{\theta}} \tag{13}$$

The share of online consumers is decided by the relative channel preference $s(\omega)$ and the shape parameter θ[24]. Similar to Redding and Weinstein (2016), the expenditure share for each variety ω and each consumer μ is:

$$f_j(\mu, \omega) = (\frac{p(\mu, \omega)/z(\mu, \omega)}{P_j})^{1-\sigma} \tag{14}$$

The online expenditure share of consumer μ across varieties can be expressed as $\int_{\Omega_j} f_j(\mu, \omega) \pi_{oj}(\omega) d\omega$, and the online expenditure share in a city is,

$$\pi_{oj} = \int_{\Omega_j} \int_{U_j} f_j(\mu, \omega) \pi_{oj}(\omega) d\mu d\omega \tag{15}$$

Note that because the expenditure share for ω for consumer μ depends on the price index P_j of the city where consumer μ resides, and P_j depends on shipping cost τ_{ij}, wage w_j and productivity ϕ_j, the shares of online shopping π_{oj} are not the same across cities. In the absence of the data on city characteristics, and given the intention to focus on the analysis of switching shopping channels, I assume that consumers have the same relative preference for all products, $s(\omega) = s$ for any ω. I also assume that all firms can serve all cities in both channels, that is, $\Omega_j = \Omega_i$ for any $i, j \in I$. As a result of these assumptions on symmetry, the model predicts that the shares of online shopping across cities are the same[25]. I discuss the data limitation in allowing for a heterogenous s in Section 2.3.

2.1.5 Price Shock and Channel Substitution

Due to the symmetry assumptions for s across varieties, I focus on a specific variety ω produced in city i and sold in j. I suppress index ω for simplicity. One of the key findings in Eaton and Kortum (2002) (EK model) is that the price distribution of the varieties that any given origin actually sends to any given destination is independent of its origin regions. In my model, firms do not compete in price given the monopolistic competition assumption, and the listing price is not a random variable as the listing price is based on the expectation of aggregate demand. Instead, the quantity demanded is a random variable as matching quality is a random variable. Similar to the EK model, the distribution of quantity demanded across consumers in city j through a channel $m \in \{o, f\}$ is

---

[24]These fractions are similar to those in Dolfen et al. (2019). In their model, the relative preference for online shopping s is conceptualized as two factors: relative quality of online merchants and ease of access.

[25]However, the model allows firms in different cities to be different in productivity, wage, shipping cost, which implies that consumers in different cities allocate their income differently. Classic predictions from many trade model hold here. For example, remote cities have higher price index due to higher shipping cost assuming all else being equal.

independent of channels[26],

$$Pr\{q_o(\mu) \leq \tilde{q}|q_o(\mu) \geq q_f(\mu)\} = Pr\{q_f(\mu) \leq \tilde{q}|q_f(\mu) \geq q_o(\mu)\} = G_j(\tilde{q}) \qquad (16)$$

Intuitively, what is happening is that the channel with better consumer appeal (higher s_m) can serve a greater number of consumers exactly up to the point where the distribution of quantities for what it sells through channel m is the same as m's overall quantity distribution. For example, if offline grocery shopping in Waitrose[27] is more convenient than online grocery shopping on Amazon, then Waitrose has a larger consumer base than Amazon, to the point at which the quantity served by Waitrose will have the same distribution as the quantity that consumers shopped on Amazon[28]. Because the distributions of quantities are the same in the two channels, the expected quantity will be the same in the two channels.

$$E(q_o|q_o(\mu) \geq q_f(\mu)) = E(q_f|q_o(\mu) \leq q_f(\mu)) = E(q) \qquad (17)$$

Using to the example above, this property implies that although there are more offline Waitrose consumers relative to online Amazon consumers (assuming going to Waitrose stores is more appealing than waiting for deliveries from Amazon), an online consumer does the same amount of grocery shopping on Amazon as an offline consumer does in Waitrose. These predictions derived from the assumption that matching quality follows the Fréchet distribution appear remarkably plausible. Given the total quantity consumed in channel m equals the number of consumers that choose m multiplied by the average quantity consumed in channel m, equation 17 implies that the quantity consumed through channel m is,

$$Q_{mij} = \pi_{mj}L_j E(q_m|q_m(\mu) \geq q_{n \neq m, n \in \{o,f\}}(\mu)) = \pi_{mj}L_j E(q) = \pi_{mj}Q_{ij} \qquad (18)$$

Equation 18 shows that the expected quantity consumed through channel m equals the share of consumers that choose channel m multiplied by the expected overall demand from both channels[29]. Importantly, these properties hold when prices change. The change in channel-specific demand can be approximated by a derivative formula. During the online shopping event, price shock k

---

[26]Another way to think about this is, the perceived price, or the price adjusted by the matching quality $\frac{p}{z_m}$, is a random variable. Thinking z as the productivities draw in the EK model, then the perceived price is the same as the purchase price in the EK model. Quantity is a function of the perceived price, so it is independent of channels (origin region in the EK model).

[27]Waitrose is a chain of British supermarkets, similar to Whole Food in the US.

[28]If we think the share of offline consumers for a product as the frequency of offline shopping for a particular consumer, then this property implies that although the consumer does offline shopping more frequently assuming offline shopping is more convenient (higher s_f), the distribution of the number of goods the consumer purchased does not depends on the channel.

[29]Note that this property relies heavily on the Fréchet distribution assumption.

increases from 1 to a higher value. Taking derivatives of Q_{mij} on k gives

$$\frac{dQ_{mij}}{dk} = \underbrace{\pi_{mj}\frac{dC_{ij}A}{dk}}_{\text{mean effect}} + \underbrace{C_{ij}A\frac{d\pi_m}{dk}}_{\text{share effect}} \tag{19}$$

I denote the first term as the mean effect and the second term as the share effect. Three interesting results emerge. First, the ratio of the mean effect to the share effect of the offline channel is constant $-\frac{\sigma-1}{\theta}$. Second, the ratio of the share effect in the online channel to the share effect in the offline channel is -1. Third, the ratio of the mean effect to the share effect of the online channel is,

$$\gamma = \frac{\sigma-1}{\theta}(k^{\frac{\sigma}{\sigma-1}}s)^\theta$$

which is an increasing function of s and k. Specifically, $\gamma = \frac{\sigma-1}{\theta}$ when $s = 1$ and $k = 1$. In this case, the online and offline channels are symmetric in terms of preference and price differences. Above linkages between these terms allow me to express these effects by one of these effects. I choose to normalize these effects based on the share effect of the online channel. Define ν as the share effect of the online channel. The other parts can be shown as below,

$$\frac{dQ_{oij}}{dk} = (\underbrace{\pi_o\frac{dA}{dk}}_{\gamma\nu} + \underbrace{A\frac{d\pi_o}{dk}}_{\nu})C_{ij} \tag{20}$$

$$\frac{dQ_{fij}}{dk} = (\underbrace{\pi_f\frac{dA}{dk}}_{\frac{\sigma-1}{\theta}\nu} + \underbrace{A\frac{d\pi_f}{dk}}_{-\nu})C_{ij} \tag{21}$$

Consumption in city j can then be obtained by aggregating through origin cities $i \in I$. Therefore, the online-offline substitution ratio λ_j, defined by the ratio of the change in online shopping to the change in offline shopping in city j, is,

$$\lambda_j = \frac{\sum_i \Delta Q_{oij}}{\sum_i \Delta Q_{fij}} = -\frac{\theta + (\sigma-1)(k^{\frac{\sigma}{\sigma-1}}s)^\theta}{\theta - \sigma + 1} \equiv \lambda \tag{22}$$

The online-offline substitution ratio is the ultimate goal of the demand model. It has following properties:

(i) This ratio is the same across cities due to the symmetry assumption of s.

(ii) Given the assumption $\theta > \sigma - 1$, it follows that $\lambda < 0$, which guarantees that offline consumption decreases when online shopping increases.

14

(iii) As along as $\sigma > 1$, $|\lambda|$ is always greater than 1. If varieties are substitutable, then $\sigma > 1$ is satisfied. This indicates that the switch from offline shopping to online shopping is not one-to-one. In the trade literature, σ is frequently estimated to be greater than one.

(iv) The inverse of the absolute value of the ratio $\frac{1}{|\lambda|}$ determines the amount of offline shopping that consumers are willing to substitute with a one unit increase in online shopping, and therefore dictates the amount of traffic related to offline shopping that can be saved.

Especially, it turns out that λ can be conveniently estimated using the elasticity of online consumption quantity to the relative price of online to offline channel,

$$\frac{1}{|\lambda|} = 1 - \frac{\sigma}{\rho} \tag{23}$$

where ρ is the elasticity of online consumption quantity to the relative price of online to offline channel. To be brief, this follows from equation 12 and equation 20, by expressing s^θ using the information on the share of online shopping π_{oj},

$$\rho = \frac{(\sigma - 1)s^\theta + \theta}{s^\theta + 1} \frac{\sigma}{\sigma - 1} \tag{24}$$

where $\rho \approx \frac{\widetilde{Q}_{oj}}{\Delta k}$ [30]. $\widetilde{Q}_{oj} = \frac{\Delta Q_{oj}}{Q_{oj}}$ denotes the growth rate of online shopping quantity. Δk is the change in the relative price shock. This holds when Δk is small[31]. ρ can be calculated using data on $\widetilde{Q}_{oj}$ and Δk. See Appendix C.0.1 for a derivation.

Equation 23 shows that $\frac{1}{|\lambda|}$ increases with the elasticity of online consumption quantity to the relative price of online to offline channel ρ, while decreases with the elasticity of substitution between varieties σ. Intuitively, if consumers are more sensitive to the price difference between channels (a higher ρ), consumers reduce more offline shopping with one unit increase in online shopping; if products are highly substitutable (a higher σ), then consumers reduce less offline shopping with one unit increase in online shopping. Note that σ measures the substitutability across products while λ measures the substitutability across channels. As certain approximations are involved in the derivation, Appendix C.0.2 provides a simulation to validate the formula for λ.

---

[30] ρ_j could vary across cities as s_j is potentially heterogenous and vary across cities. As s_j is assumed to be the same across products and thus across cities, ρ_j reduces to a scalar ρ.

[31] When Δk is small, $\frac{\widetilde{Q}_{oj}}{\Delta k} = \frac{\frac{\Delta Q_{oj}}{Q_{oj}}}{\frac{\Delta k}{k}}$ approximates the elasticity of online consumption quantity to the relative price of online to offline channel. $k = 1$ due to the assumption of equal price across channels. Note that k increases when the online price decreases in the set-up of my model, so ρ is positive and is the absolute value of the elasticity of online consumption quantity to the relative price of online to offline channel. Denote $p = \frac{1}{k}$ as the relative price, $\frac{\Delta k}{k} = p\Delta\frac{1}{p} \approx -\frac{\Delta p}{p}$.

The simulated λ is slightly smaller than its theoretical value in equation 22 due to an omitted higher order component in equation 19, but are almost identical to the value obtained using equation 23.

2.2 Traffic Congestion and Online Consumption

Assuming the two ways of shopping have different levels of traffic efficiency, the overall traffic due to shopping will then change during the event. The following section combines insights from transportation engineering literature and accounting assumptions on vehicle demand and traffic density to derive the elasticity of traffic congestion index to the quantity of online shopping. The elasticity includes three components: traffic density, the importance of e-commerce, and a traffic-saving factor, which is per-unit online good traffic saving. Note that the model ignores the effect of traffic congestion on vehicle demand given the unique context of the event, for which I will provide more details.

2.2.1 Online Consumption and Vehicle Demand

Given a time interval in city j, shopping vehicle demand D_j can be calculated as the sum of online shopping vehicle demand and offline shopping vehicle demand,

$$
\begin{aligned}
D_j &= t_o \sum_i Q_{oij} + \zeta_j t_f \sum_i Q_{fij} \\
&= t_f \delta Q_{oj} + \zeta_j t_f \frac{\pi_f}{\pi_o} Q_{oj}
\end{aligned}
\tag{25}
$$

where t_o is the per-unit good vehicle demand for online shopping and t_f is the per-unit good vehicle demand for offline shopping. $\delta = \frac{t_o}{t_f}$, which I call vehicle-saving ratio. It is smaller than one if online shopping is more vehicle-efficient relative to offline shopping. ζ_j is the average share of shopping made through private vehicles[32]. Thus, $\zeta_j t_f \sum_i Q_{fij}$ is the vehicle demand for offline shopping. Now, denoting the share of shopping vehicle demand to the overall vehicle demand (including vehicles for other purposes such as commute) on the road as ψ_j, then the overall vehicle demand in a city is $\frac{D_j}{\psi_j}$. Denoting the capacity of roads (for example, the total length of roads) in the city as R_j, then the traffic density on the roads (vehicle/km),

$$
n_j = \frac{D_j/\psi_j}{R_j}
\tag{26}
$$

---

[32]Bus travel can be discounted into car travels, which is ignored here for simplicity. Other forms of transporting shopping goods include walking or taking public subways.

Given the change of online consumption ΔQ_{oj}, the change in shopping vehicle demand is the sum of the change of vehicle demand in online shopping and that in offline shopping.

$$\Delta D_j = t_o \sum_i \Delta Q_{oij} + \zeta_j t_f \sum_i \Delta Q_{fij}$$
$$= t_f (\delta - \frac{\zeta_j}{|\lambda|}) \Delta Q_{oj} \qquad (27)$$

Similarly, the change of traffic density is

$$\Delta n_j = \frac{\Delta D_j}{R_j} \qquad (28)$$

Combining equation 25, 26, 27, 28 gives,

$$\frac{\Delta n_j}{n_j} = \frac{\psi_j (\delta - \frac{\zeta_j}{|\lambda|})}{(\delta + \zeta_j \frac{\pi_f}{\pi_o})} \frac{\Delta Q_{oj}}{Q_{oj}} \qquad (29)$$

Equation 29 presents the relationship between the changes in traffic density with the changes in online shopping.

Before moving on to introducing the relationship between traffic congestion index and online shopping, it is worth noting that the change of shopping vehicle demand for intercity logistic is,

$$\Delta D_{ij} = t_h \Delta Q_{oij} \qquad \text{if } i \neq j \qquad (30)$$

where t_h is the per unit of good vehicle demand for online shopping on the intercity roads. This equation simply states that the increase in online shopping increases vehicle demand for intercity travel owning to offline retail logistics. Of course, the reduction in offline shopping may reduce intercity traffic; however, this negative adjustment may materialize much slower than the sharp increase in the demand for online shopping. For this reason, I assume that possibility away.

2.2.2 Vehicle Demand, Traffic Density, and Traffic Congestion Index

Now the task is to provide a mapping from vehicle demand to traffic congestion index using traffic density. Note that the model assumes that vehicle demand increases traffic density, and thus increases traffic congestion, while ignores the effect of traffic congestion on vehicle demand. The reason is that empirical evidence shows that the change in traffic congestion is very small. Traffic congestion index reduces by 4%, which is about 2 minutes time reduction for a one hour travel. Such small change is arguably undetectable by commuters, at least in the one week post-event time

window that this research focuses on. Therefore, I model a one-way relationship between vehicle demand and traffic congestion.

As summarized by Yang et al. (2018a), traffic speed and density follow a monotonic relationship. Density further reflects vehicle demand monotonically because a decision to use road transport is essentially a decision to add a vehicle on the road (Else, 1981; Walters, 1961). I follow the functional form of speed and density as in Adler et al. (2017)[33] and derive the relationship between the traffic congestion index and density,

$$lnT_j = \frac{n_j}{n_{mj}} - ln(u) \tag{31}$$

where T_j is the traffic congestion index for a given road segment. n_{mj} is the density of vehicles on the roads when maximum flow is achieved, and u is a constant[34]. Using equation 29, the marginal change in traffic congestion index can be expressed as

$$\Delta lnT_j = \frac{\Delta n_j}{n_{mj}} = \frac{n_j}{n_{mj}} \frac{\psi_j \pi_o}{\delta \pi_o + \zeta_j \pi_f} (\delta - \frac{\zeta_j}{|\lambda|}) \frac{\Delta Q_{oj}}{Q_{oj}} \tag{32}$$

When the change of ΔQ_{oj} is very small, $\frac{\Delta Q_{oj}}{Q_{oj}} \approx \Delta lnQ_{oj}$[35]. Hence, I derived the elasticity of traffic congestion to online consumption,

$$\epsilon_j \approx \frac{n_j}{n_{mj}} \frac{\psi_j \pi_o}{\delta \pi_o + \zeta_j \pi_f} (\delta - \frac{\zeta_j}{|\lambda|}) \tag{33}$$

ϵ_j is a variable that varies across cities. ϵ_j can be decomposed into three parts: the first part $\frac{n_j}{n_{mj}}$ is the ratio of actual density to the optimal density, which measures the congestion level. The intuition for this term to appear in the equation is that the impact of the reduction of vehicles is stronger when roads are more congested. The second part $\frac{\psi_j \pi_o}{\delta \pi_o + \zeta_j \pi_f}$ captures the importance of online shopping relative to offline shopping in the city. Intuitively, a higher share of online shopping in a city implies a larger scope for e-commerce to impact the city's traffic congestion. The third

---

[33]This functional form was proposed by Underwood (1961). See Brilon and Lohoff (2011) for how well this functional form fits with real-world data and other possible functional forms.

[34]Omitting city index j, Underwood density-speed equation is

$$n = n_m ln(\frac{v_r}{v})$$

where v is speed, and v_r is the "reference" speed, which is estimated to be 300km/h for typical motorways conditions in the transport engineering literature. Assuming that the reference speed is u times to the free-flow speed v_f, traffic congestion index $T = \frac{time}{time_f} = \frac{1/v}{1/v_f} = \frac{1/v}{u/v_r} = \frac{1}{u} e^{n/n_m}$ where $time$ is the actual passing time of a road segment and $time_f$ is the free-flow passing time. See Notley et al. (2009) for details.

[35]It is easy to show that $\Delta lnQ = ln(1 + \frac{\Delta Q}{Q}) \approx \frac{\Delta Q}{Q}$ when ΔQ is very small.

18

part $\delta - \frac{\zeta_j}{|\lambda|}$ determines the traffic saved by per-unit online good, which I term "traffic-saving factor". Importantly, as the first two parts of ϵ_j are always positive, the condition for online shopping to reduce traffic congestion is simply,

$$\delta < \frac{\zeta_j}{|\lambda|} \tag{34}$$

The intuition of the condition is that, for the increase in online shopping to result in a reduction in vehicles, the vehicle-saving ratio is sufficiently low, or the amount of offline shopping that consumers are willing to substitute with online shopping is sufficiently large.

In the following sections, I first conduct a quantitative analysis of the elasticity ϵ_j, then focus on estimating the mean of the elasticity $\epsilon = \bar{\epsilon}_j$ empirically. Note, I use equation 32 as the estimation equation for ϵ instead of equation 33 because ΔQ is large[36].

2.3   Quantitative Analysis of the Model

In order to quantify the elasticity of traffic congestion to online shopping using the model, I begin by assuming values of parameters or sample means of some variables for the three components in equation 33. Table 1 lists the parameter values and sample statistics used to estimate the elasticity in equation 33. I first estimate λ and the vehicle-saving ratio for per unit of good, with reasonable assumptions of σ and using the sample moment of ρ. Note that there are some complications in transferring the spike in the Baidu Index to the growth rate of online shopping quantity between the two weeks surrounding the event[37]. A very crude procedure is to use the average daily online consumption (estimated using yearly online consumption) and consumption on the event day (estimated using reported overall sales on the event day) to calculate the weekly online consumption growth rate. The extra online consumption due to the event in the first week can be evenly added to the normal online consumption stream in the second week. If the numbers from the two sources do not align, I can adjust the growth rate of the Baidu Index, accordingly, to better approximate the weekly growth rate of online shopping. The weekly growth of consumption is estimated at about 160%, which turns out to coincide with the sample mean of the growth rate of the Baidu Index. For this reason, I just use the growth rate observed from the Baidu Index without any adjustments. Admittedly, estimation errors may arise due to a lack of information. Because the online prices are reported to be about 80% of the online price on the event day as mentioned earlier, the change in price shock is set to 0.25 (= $\frac{1}{0.8} - 1$). The estimate of ρ is then 6.4 (= 1.6/0.25). The elasticity of substitution between varieties σ is assumed equal to four, which is in line with figures frequently used in the international trade literature (see Redding and Sturm (2008)). Using

---

[36]The estimation of the elasticity ϵ based on equation 32 is therefore an approximated value.

[37]For simplicity, I assume that the impact of the shopping event only lasts for a week.

equation 23, the online-offline substitution ratio λ is estimated to be -1.9, which suggests that offline shopping reduces a half unit when online shopping increases one unit. Because a fraction ζ of offline consumption is made by using private vehicles or taking taxis, the decrease in offline shopping vehicle demand is $\frac{\zeta}{1.9}$ units assuming that the offline consumption reduces proportionally for consumers who use private vehicles and for consumers who do not use private vehicles. ζ is estimated to be 0.91, derived by calculating the share of shopping trips in private vehicles or taxies to the total shopping trips during peak hours in urban areas in the United States using the NHTS 2017 data[38]. It is reported[39] that a typical Amazon driver can delivery about 150-200 parcels a day (about 10 hours, according to the report.). Given that it takes consumers about one hour for a round trip for retail shopping[40], and assuming the amount of good that consumers buy is equivalent to a parcel in the trip, then online delivery is about 15 times more efficient than offline retail[41]. Thus, δ is assumed to be 0.067. For a one unit increase in online shopping, about 0.48 (= 0.91/1.9) units of vehicles used for offline shopping are saved, while additional online shopping vehicles is only 0.067 unit of vehicles. Taken together, the vehicle-saving ratio per unit online shopping quantity is -0.413(=0.067-0.48)[42].

Next, I set the values of parameters or estimate sample statistics in $\frac{\psi \pi_o}{\delta \pi_o + \zeta \pi_f}$. π_o is estimated equal to be 12.6% according to national statistics on the share of online sales to overall retail sales. I do not have city-level data on the share of online consumption, which is the reason for assuming homogeneity in $s(\omega)$. The share of online shopping grows dramatically since online shopping increases by 160% on average in the week after the online shopping event. I postulate the value of the share of online shopping in the week after the event using its initial value and the growth rate of online shopping, and then take the mean of both values for calculating the elasticity. ψ is estimated based on the US NHTS 2017 data as well. I calculate the share of shopping trips using vehicles to total trips using vehicles during peak hours in urban areas. Finally, I estimate the value for $\frac{n}{n_m}$. This part can be expressed as $\frac{n}{n_m} = ln \frac{v_f u}{v} = lnu + lnT$. lnT can be estimated using the sample mean of the traffic congestion index, which is 1.65 in peak hours in the two weeks surrounding the event. u is estimated using the ratio of the reference speed, which I set to 300 km/h, and free-flow speed, which I set to 60 km/h. Collectively, the elasticity of traffic congestion index to online shopping

---

[38]The reason to use the US data is that I do not have access to Chinese household traffic survey data. The figure observed from the US data is likely to be larger than that in Chinese cities as the number of road vehicles per capita is much higher in the US. See `https://en.wikipedia.org/wiki/List_of_countries_by_vehicles_per_capita`

[39]See `https://www.bbc.co.uk/news/uk-england-37912858`.

[40]According to the US NHTS 2017 data, an average shopping trip using cars take 27 minutes one-way.

[41]Amazon driver delivers about 15(=150/10) parcels in an hour, while a consumer buys a parcel in an hour.

[42]Note I have only considered vehicle savings from the change of logistics from consumers to stores. There may also be vehicle savings from the change of logistics from warehouses to stores in the long term, which is, however, unlikely to be an issue for this study as I focus on a short term sales event.

quantity is estimated to be −0.06. However, the estimates based on the quantitative model vary substantially when the share of online shopping and the growth rate of online shopping change. If increasing the values of the share of online shopping to 0.2, which is the statistic for Beijing in 2016, the elasticity is estimated to be −0.11. If I also increase the values of the growth rate to the maximum value observed in the sample (2.7), then the elasticity is −0.25. Therefore, while the model predicts a negative elasticity of traffic congestion to online consumption quantity, it does not identify exact magnitudes. For this reason, I turn to empirical estimates.

3 Data and Descriptives

3.1 Data on Traffic Congestion and Pollution

This study collected traffic congestion data from a GPS navigation company in China. The traffic congestion index is the ratio of the actual passing time to the free-flow passing time for a given road segment recorded from the company's millions of GPS navigation service users. The data contains 94 major Chinese cities. An average city in the sample has a population of 3 million[43] and an average annual GDP per capita of 86 thousands *yuan*.

Figure 8 shows the time series of traffic congestion index in the period that is close to the shopping event in 2016. The green line shows daily average traffic congestion; the red line shows traffic congestion during peak hours from 7 am to 9 am and from 5 pm to 7 pm; the blue line shows traffic congestion during off-peak hours. The unit in the horizontal axis is the number of days away from the online shopping event on 11 November 2016. The big drop of these lines around day 77 shows the Chinese New Year holiday, when people enjoy the holiday and commute much less[44]. The days between the vertical lines in the left of the graph are the weeks surrounding the online shopping event of 11 November and an offline shopping event on 12 December in 2016. Traffic congestion levels start high on Monday, drop in the middle of the week, and bounce back to another peak on Friday, before plummeting over the weekend. Given the high volatility in the time series of congestion data, I restrict the sample to a narrow band around the event: weekdays of one week before the event and two weeks after the event. This restriction reduces the unobserved time trends of traffic in the study period and highlights the impact of the online shopping event on traffic congestion[45].

---

[43]The population data is based on household (Hukou) registration, which does not include migration population. The number is usually smaller than the actual population in cities.

[44]Many migrant workers return to their hometowns, often in rural areas. Therefore, the population temporarily drops in cities.

[45]Ideally, I would like to have the traffic congestion index in two weeks before the online shopping event to study the pre-event trend carefully; however, I started to code the program to track the hourly update of traffic congestion index

Inspired by Akbar et al. (2018), I utilized Baidu Map API to query hourly travel time based on real-time road traffic condition across the cities in my sample in one week before and two weeks after the Singles' Day shopping Event in November 2018. There are 9,216 records of travel time information recorded from 7 am to 11 pm for each day. Routes between cities remain the same. The lengths of routes are known and remain constant. This dataset shows how the Singles' Day shopping event shifts intercity traffic pattern. The data were collected from 5 November to 23 November. I was unable to collect the data for some hours of the day. The data are mostly complete on the Thursday and Friday in the week before the event and in the two weeks after the event. In the related graphs and regressions, I added hour × weekday fixed effect to address the missing variable issue. In the regressions, I restrict sample size to Thursday and Friday only.

Air pollution data are published online hourly by China's Ministry of Environmental Protection. I collected the data for about 1,563 monitoring stations across 337 prefectures. The measures include air quality index (AQI), carbon monoxide (CO), nitrogen dioxide (NO2),ozone (O3), particulate matter with a diameter of 2.5 μm or less (PM2.5), particulate matter with a diameter of 10 μm or less (PM10), and sulfur dioxide (SO2). NO2 concentration is the primary outcome variable of interest as it is a major vehicle exhaust. Figure 11 shows the daily cycle of the shock of NO2 and traffic congestion index[46]. The timing of the generation of NO2 in each hour follows the change of traffic congestion index closely in the daytime. I further plot daily NO2 level and traffic congestion index in a longer time horizon as shown in Figure 12. NO2 follows traffic congestion index closely across days, especially when traffic congestion index plummeted during important holidays such as the Chinese New Year Festival. NO2 concentration is a good predictor of traffic congestion. Other pollutants do not correlate with traffic congestion as well as NO2.

Table 2 reports summary statistics for our dataset in the two weeks window surrounding the Singles' Day Shopping Event. Each observation is city-by-hour. The first two columns report the outcomes of interest one week before and one week after the event for the peak hours, and the last two columns report those for off-peak hours. The average traffic congestion index for peak hours is 1.74 before the event, and dropped to 1.67 after the event. The hourly concentration of pollution appears to be much more volatile than traffic congestion. Pollution levels increase substantially in the week after the shopping event; however, its increase does not correlate with the change of online shopping as shown later in Section 6.5.

---

on the website of navigation company on 6 November 2016, and the website does not provide data retrospectively. The data collection was also interrupted during the first weekend after the event due to technical reasons, so I cannot study the effects over the weekends.

[46]I estimated hourly shock of NO2 following procedures in Henderson (1996).

3.2 Data on Online Shopping

To measure the change in online shopping, I use the frequency of searching the name of the online shopping platform in each city based on Baidu Index. Baidu is the largest search engine service provider in China and dominates the market since Google search engine exited the Chinese market in the year 2010. Baidu Index is a publicly available web service and bears a number of similarities to Google Trends. The query index is based on the frequency of the search keywords within a day as the minimal unit. Importantly, it provides the IP addresses of its users to city level. This feature allows tracking the trends of search frequency of the online shopping platform for each city. Web search engine index has been previously adopted to track economic activities in real time (Choi and Varian, 2012). Vosen and Schmidt (2011) show that forecasting of monthly private consumption based on Google Trends outperforms survey-based indicators.

Figure 1 depicts an example output from Baidu Index for the query of the two possible names of the online shopping platform: Taobao and Tmall. Taobao is a consumer-to-consumer (C2C) online retail platform for small businesses and individual entrepreneurs to open online stores. In contrast, Tmall runs a business-to-consumer (B2C) online platform for local Chinese and international businesses to sell brand name goods. Tmall shops are required to have established physical stores, and often have national offline distribution channels. In short, Taobao operates like eBay while Tmall operates like Amazon. Unlike either eBay or Amazon, Alibaba provides both services and the product searches in either platform give results from both Taobao and Tmall[47]. Assuming that a constant share of the population entering the online shopping website through the search engine in cities, the index can serve as a proxy of the number of online consumers and captures the increase of online shopping during the shopping event day. As shown in Figure 1, daily searches on the online shopping platform escalated from around 900 thousand times before the event to around 2.7 million times at the peak on 11 November 2016. I extracted the index for the day of 11 November 2016 which is the event day, and 11 October 2016 which is one month before the event. I only used the index on the event day instead of a cumulative sum of the index around the event, given the fact that the sale only lasts for one day and consumers have to complete the order on the day of the event[48]. The daily value in a month before the event shows the event-free frequency of visiting the online platform. I choose 11 October 2016 to represent the event-free average search as it is far enough from the event in time but not too far, and using the same day as the event day in the last

---

[47]Consumers can easily search goods from the two sources in the same search entry box in either of the two domains of the platform. Therefore, I use the sum of the search of both platforms in Baidu Index instead of only using Tmall, despite the fact that the online shopping event is only for Tmall stores.

[48]The index is above average in the few days around the event day. Consumers may browse products before the event and check the status of delivery after the event. Using the cumulative sum of these searches may double count the actual number of transactions.

month could avoid possible bias in the seasonal trend of searches within a month[49].

The online shopping platform also publishes an index that measures e-commerce development in cities: Alibaba E-commerce Development Index (AEDI). AEDI is the weighted average of an online shopping index and an online selling index. The online shopping index is a weighted average of the number of online buyers and average online consumption; The online selling index is a weighted average of the number of online sellers and average online sale. The two indices thus measure the intensity of online shopping and online selling in each city, respectively. Equations in Appendix B show how the two indices are constructed. The Baidu index and AEDI cover about 277 cities while the traffic congestion index only covers 94 major cities. The overlap cities form the main sample of the study. I occasionally use the expanded sample with 277 cities when I do not need the traffic congestion index.

To validate the e-commerce indices, I collect actual online consumption or online sales data from news and online reports. In Figure 9(a), I record the first-hour online consumption data by provinces in the Singles' Day event in 2016. I took the average of the city-level online shopping index to create an index at the province level. The graph shows that log online consumption is positively correlated with online shopping index[50]. In Figure 9(b), I collected monthly online sales data of Alibaba in May 2017, by cities. Plotting it against the online selling index shows that the log online selling index is a good predictor of log online sales in cities.

I extracted city-to-city postage fee per km from the website of a leading national logistics company and derived the average postage fee for each city[51]. Specifically, I take the simple average of postage fee for a destination city across all origin cities. I calculate the distance matrix of cities based on their centroids. Coupled with population data, I measure market access for each city. In addition to the overall trend in online shopping measured by the Baidu Index, and the cross-sectional variation measured by the online shopping indices, I obtain the number of online stores in different categories for both Tmall and Taobao in each city. I assign the online stores to their registration cities, despite that being online means that it can provide its products to all cities. This data allows me to examine the heterogeneity in the effects of online shopping on traffic congestion.

---

[49]It is possible to obtain an average of Baidu Index over a period; however, there is a daily cap for querying these indices, which makes obtaining more data time consuming and difficult. As shown in 1, the index was very flat before the event. Obtaining more days is unlikely to change the pre-event average of the index.

[50]The relationship holds when controlling for GDP and population.

[51]Logistics is a highly competitive industry in China, so the price should be very similar across firms

4 Econometric Models for Online Shopping and Traffic Congestion

This section discusses the econometric models to estimate the effect of online shopping on traffic congestion. First, I present simple regression models that quantify the changes in intracity traffic congestion index and intercity travel time surrounding the event. Then, I present the ordinary least squares (OLS) estimates and instrumental variable estimates of equation 32. Third, I present an event-study approach as a robustness check. Due to reasons mentioned in Section 3.1, I restrict the timeframe of the analysis to weekdays in one week before the online shopping event and one or two weeks after the online shopping event. Note that all mathematical notations in the regression models have different meanings compared to those in the theoretical model section unless explicitly specified otherwise[52].

4.1 Quantifying the Changes in Travel Time and Traffic Congestion Surrounding the Event

To quantify the change of intracity traffic congestion, I estimate equation 35,

$$lnT_{iht} = \sum_t \gamma_t Week_t + \sum_t \beta_t Week_t lnO_i + \sum_t \delta_t Week_t lnS_i + \sum_t Week_t X_i \zeta_t + \iota_i + \lambda_h + \psi_w + \varepsilon_{iht} \quad (35)$$

where t indexes weeks, with $t = 0$ indexes the week before the event and $t = 1, 2$ indexes the first and second week after the event. The dependent variable lnT_{iht} is the log traffic congestion index in city i at hour h (in week t). $Week_t$ are dummies indicating the first and second week after the shopping event, with the reference group being the week before the shopping event. ι_i is city fixed effect, λ_h is the hour of the day dummy, and ψ_w is the day of week dummy. γ_1 and γ_2 capture the average change of the traffic congestion index in the first and second week after the shopping event, respectively. I further expect $|\gamma_2| < |\gamma_1|$ as the impact of the event would fade away. To relate the change of traffic congestion to heterogeneous shocks experienced by cities during the online shopping event, I include the interaction of log online shopping index O_i and log online selling index S_i with the week dummies in the regression model. X_i are control variables such as income and the number of internet and mobile users in cities. The interactions of X_i with week dummies control for potential city-specific trends that correlate with income and the number of online consumers.

---

[52]This relieves the burden of finding new Greek letters.

Similarly, I quantify the change of intercity travel time with equation 36,

$$Time_{ijht} = \alpha Distance_{ij} + \sum_t \gamma_t Week_t + \sum_t \beta_t Week_t lnO_i + \sum_t \delta_t Week_t lnS_i$$

$$+ \sum_t Week_t X_i \zeta_t + \iota_i + \kappa_j + \lambda_h + \psi_w + \varepsilon_{ijht} \tag{36}$$

where $Time_{ijht}$ is the intercity travel time from city i to j at hour h (in week t). $Distance_{ij}$ is the length of each route between city i and j in km. ι_i and κ_j are origin and destination city fixed effects. As predicted by equation 30, roads between cities are expected to be filled with trucks that deliver goods from manufacturers or warehouses to distributors, therefore I expect at least $\gamma_1 > 0$ and $\gamma_1 > \gamma_2$ as the effects are expected to weaken over time.

4.2 Ordinary Least Square Estimation of the Effect of Online Shopping on Traffic Congestion

This section explores the variation in the increase of online shopping index across cities to estimate the relationship between online shopping and traffic congestion. Estimating equation 32 derived from the theory section requires estimating below first-differences regression model,

$$\Delta lnT_i = \beta \tilde{B}_{it} + \varepsilon_i \tag{37}$$

where i indexes cities. ΔlnT_i is the change of log traffic congestion index. $\tilde{B}_{it}$ denotes the growth rate of Baidu Index for the online shopping platform $\frac{\Delta B_i}{B_i}$, which is the proxy for $\frac{\Delta Q_{oj}}{Q_{oj}}$ in equation 32. β estimates the mean elasticity ϵ in the theoretical model.

There are occasional missing values in the traffic congestion index. If the missing values happen on hours or day of week with particularly high or low traffic congestion, then the difference of the outcome variable between weeks may arise due to missing values. Therefore, I add hour of the day and day of week fixed effects to avoid potential biases. For these reasons, I estimate below level specification,

$$lnT_{iht} = \beta \frac{B_{it}}{B_{i0}} + \lambda_h + \psi_w + \mu_{iht} \tag{38}$$

where $t = 0, 1$ with $t = 0$ indicating the week before the event and $t = 1$ indicating the week after the event. The dependent variable lnT_{ih} is log traffic congestion index in city i at hour h. B_{it} is the value of Baidu Index in time t, and B_{i0} is the Baidu Index for the week before the event. In $t = 0$, the value of $\frac{B_{it}}{B_{i0}}$ is always 1. The construction of the regressor allows the coefficient from the level specification have the same interpretation as in the change specification as $\Delta \frac{B_{it}}{B_{i0}} = \frac{\Delta B_i}{B_i}$ given there are only two periods. The value of B_{i0} is taken on the day of 11 Oct 2016, and B_{i1} is

measured by the peak value on the day of the event, as explained in Section 2.3. λ_h represents hour dummies, and ψ_w represents the day of week dummies. The residual μ_{iht} in equation 38 can be further decomposed into three components,

$$\mu_{iht} = \iota_i + \tau_t + \varepsilon_{iht}$$

where ι_i represent city-specific unobserved components that are fixed over time. τ_t represents general time effects due to the seasonality in traffic congestion, and error term ε_{iht}. I include city fixed effects to account for ι_i and time trends to account for τ_t in the regression model.

There are three potential issues with OLS estimation of equation 38. The first is measurement error. I measure the change of online shopping by the number of searches of the online shopping platform. If the measurement error is "white noise", then it biases the OLS estimates toward zero, which is the well-known attenuation bias[53]. The second is reverse causality. The estimate of β will be biased if the change in traffic congestion can affect the change of online shopping. Consumers in the cities with a higher level of traffic congestion may prefer online shopping. The city fixed effects can alleviate this concern of the effect of the level of traffic congestion on the level of online shopping since I essentially regress the percentage increase in traffic congestion on the percentage increase in online shopping. However, the change in traffic congestion might affect the change in online shopping. Consumers that observed a higher reduction in traffic congestion have a higher chance to choose offline shopping or other types of travels by cars, which leads to an overestimation of the effect (or an underestimation of the absolute value of the effect if it is negative). As argued in Section 2.2.2, the temporary reduction in traffic congestion is unlikely to be detected by commuters, especially as the study limits the timeframe to a narrow time band[54]; however, this cannot rule out the risk completely. The third issue is omitted variable bias (OVB). Although the city fixed effects and common time trend have eliminated the possible correlation between the level of online shopping with the residual, the change of online shopping may correlate with the city-specific trend in the residual, that is, $\Delta\varepsilon_i$ correlates with $\frac{B_{it}}{B_{i0}}$. For example, the true model may include the interaction terms of time trends τ_{it} with road network density presumably because cities with denser road network may experience less traffic congestion under the similar level of a travel demand shock. Road network density is also likely to negatively associated with the increase in online shopping as cities with denser road network may have more street shops within walking distance of consumers, which makes online shopping a less attractive shopping

---

[53]I cannot assess whether the measurement error complies with the classic measurement error (CME) model. See Angrist and Krueger (1999) for the consequences of other types of measurement errors.

[54]Consumers have to change their travel behavior very fast in light of the change in traffic congestion, which seems unlikely. For example, Hall et al. (2019) shows that Uber drivers' earnings adjusted to fare cuts fairly slowly although the information of fare cuts is very clear given the digital platform context.

option. This leads to an overestimation of the effect. To tackle the potential endogeneity issues in the estimation above, I propose an instrumental variable (IV) identification strategy.

4.3 The Instrumental Variable Estimation of the Effect of Online Shopping on Traffic Congestion

The proposed IV is the interaction of the online event with the average postage fee between cities conditional on its market potential. The rationale behind the IV is that postage fee is a major factor in deciding the amount of online shopping in cities, and importantly the online shopping platform waived the postage fee on the day of the event. Hence, places that had higher postage fee are expected to consume more during the limited time window of free shipping as they have a higher opportunity cost for not participating in the sale event. However, postage fee is likely to correlate with other factors in deciding the trade volume of a city, which in turn may affect the change in traffic congestion during the event. First, the most significant confounding factor is the remoteness of a city in the trade network. Cities with higher postage fee are likely to locate further away from other cities. Further, the postage fee is likely to be a function of trade quantity, which is a function of remoteness. A higher trade volume means higher scale economy in the trade route, so the freight cost in each route can be reduced. Second, the importance of the size of the waived postage fee depends on the price index of a city. The model in Redding and Sturm (2008) shows that the price index is a function of market access. Cities with higher market access have lower price index. Given these considerations, I control for polynomial terms of the market potential of a city in the IV specification. The market potential is measured by the weighted sum of the population in all destinations j that can be reached from origin i by incurring transport cost c_{ij} along a specific route between i and j. That is:

$$M_i = \sum_{j \neq i} \frac{N_j}{c_{ij}} \tag{39}$$

where M_i is the market potential of city i, N_j is the population in city j, and c_{ij} is the straight line distance from city i to city j. I use the simple inverse cost weighting scheme similar to Gibbons et al. (2019) and Couture et al. (2018). For robustness checks, I construct additional sets of market access variables using other measures of market potential instead of population. One measure is the overall number of online shops in each city listed in the online shopping platform of Alibaba. Another is the number of Tmall shops in each city. Tmall shops are certified online retailers with established brands and revenues above a certain threshold. It is the Tmall shops that are available for the online shopping event during the event day.

Specifically, I estimate below system of regressions:

$$\tilde{B}_{it} = \tau^{1st} + \gamma P_i + f(M_i) + X_i\theta^{1st} + \upsilon_i \tag{40}$$

$$\Delta lnT_i = \tau^{reduced} + \delta P_i + f(M_i) + X_i\theta^{reduced} + \epsilon_i \tag{41}$$

$$\Delta lnT_i = \tau^{IV} + \beta^{IV}\hat{\tilde{B}}_{it} + f(M_i) + X_i\theta^{IV} + \varepsilon_i \tag{42}$$

where P_i is the average of waived postage fee in city i[55], $f(M_i)$ is a polynomial function of the market potential in city i, X_i represents other control variables including GDP per capita, the number of internet users and the number of mobile users. Equation 40 is the first-stage regression that estimates the effect of waived postage fee on the changes in online shopping. Equation 41 is the reduced-form regression that estimates of the effect of waived postage fee on the changes in traffic congestion. Equation 42 provides the 2SLS estimate of β^{IV}, which identifies the causal effect of the change of online shopping on the change of traffic congestion. Above IV specifications can be written in a level specification similar to equation 38. In that case, the instrument will be the interaction of the online shopping event dummy with the pre-event average postage fee[56].

The key identification assumption is that conditional on the polynomial terms of market potential and other possible controls:

1. $Cov(P_i, \tilde{B}_{it}) \neq 0$, that is, P_i affects the change of online shopping (relevance);
2. $Cov(P_i, \varepsilon_i) = 0$, that is, P_i only affects the change of traffic through online shopping (exclusionary restriction);
3. $Cov(P_i, \beta) = 0$ and $Cov(\tilde{B}_{it}, \beta) = 0$, that is, both the waived postage fee and the growth rate of Baidu Index are not correlated with the congestion relief effects of online shopping.

The validity of the first assumption can be tested in the first stage regression. The validity of the second assumption cannot be directly tested but is likely to be satisfied. Given the online shopping event is the only significant event in the short periods of two weeks, it is unlikely that the waived postage fee can affect the reduction of traffic through other intermediate factors other than the increase of online shopping. To account for the possibility that cities with different average income levels and size of consumers may respond differently to the change in postage fees, I control for the former using GDP per capita, and the latter with the number of internet users and the number of mobile users. Finally, for the third assumption, Heckman et al. (2006) show that when

---

[55]$P_i = \frac{\sum_j P_{ij}}{I}$, where P_{ij} is the postage fee from city i to j, I is the number of cities.
[56]For example, the first-stage regression model using the level of traffic congestion is:

$$lnB_{it} = \iota_i + \tau_{it}^{1st} + \gamma P_i\tau_{it}^{1st} + f(M_i)\tau_{it}^{1st} + X_i\tau_{it}^{1st}\theta^{1st} + \upsilon_{it}$$

both instrument variable and endogenous treatment variable are uncorrelated with gains from the treatment, the IV estimator can obtain the mean treatment effect (the mean of the distribution of β given the heterogeneity). It is reasonable to believe that cities do not foresee the impact of online shopping on traffic congestion and consume more online products or choose to be more responsive to the waived postage fee, because there has not been common knowledge on the gains of traffic reduction from e-commerce. This assumption implies that even if consumers adjust their travel demand based on traffic congestion simultaneously, which is unlikely to be true, the IV can provide unbiased estimates as long as the gains of traffic congestion from online shopping are unclear to commuters.

4.4 Event Study Estimates

Above specifications use traffic in the week before the online shopping event as the counterfactual traffic had the share of online shopping not changed. The counterfactual might be contaminated if consumers hold up their consumption until the event day. As a robustness check, I use traffic congestion data following other shopping events of similar influence on consumption as the counterfactual outcome. If consumers do reduce their budgets after shopping events, this approach could cancel out part of the budget reallocation effect. Specifically, I consider a follow-up event one month after the Singles' Day shopping event: the Double Twelve Shopping Event on the day of 12 December each year. On the day of the event in 2016, the event generated \$13.85 billion sales, which is along the same magnitude as the online shopping event (\$17.6 billion). Prices in the offline channel dropped significantly due to discounts in using Alipay at the counters of the stores. The magnitude of price shocks in this event is similar to that of the Singles' Day shopping event[57].

4.4.1 Traffic in the Weeks After Both Shopping Events

Denote the period dummy D_t, with $D_t = 1$ indicating the weeks after the online shopping event, and $D_t = 0$ indicating the weeks after the offline shopping event. I estimate below equation,

$$lnT_{it}^{W_t=1} = \iota_i + \theta D_t^{W_t=1} + \beta_1 lnO_i D_t^{W_t=1} + \beta_2 lnS_i D_t^{W_t=1} + \epsilon_{it}^{W_t=1} \tag{43}$$

where $W_t = 1$ indicates the weeks after both events. O_i is the online shopping index and S_i is the online selling index. I replace Baidu Index with the online shopping index because I cannot

---

[57]It used to be an online event for Taobao stores, which are mostly individual sellers like eBay merchants. As Tmall increasingly dominates the online marketplace and the Singles' Day shopping event (for Tmall) becomes exponentially far-reaching in recent years, the online impact of the Double Twelve event is reported to be negligible. In 2016, the event turned to the offline channel to promote the company's mobile payment product Alipay (similar to Apple Pay). Consumers can obtain up to 50% deals when paying using Alipay during the event.

measure the change in offline shopping during the offline shopping event, I need a more flexible way to measure the change of online and offline shopping during both events. I use the interaction of the level of online shopping before the event with the period dummy to measure the change in online shopping (replace $\frac{\Delta B_{it}}{B_{i0}}$ with $O_i D_t$), given the fact that the cities with a higher online shopping index experienced a greater increase in online shopping during the event as shown in 4. The online selling index could potentially capture the change of traffic in the cities that sell products to other cites, so I include them in the regression. β_1 is a difference-in-differences in style estimator (Cooper et al., 2011), which reflects how traffic responds differently in cities with different intensity of online shopping. If online shopping reduces traffic, I expect $\beta_1 < 0$.

4.4.2 Traffic in the Weeks Before Both Shopping Events: A Placebo Test

Replicating the above specification in the weeks before the two shopping events would serve as a placebo test. Had both events not happened, we should not observe the correlation between the change of traffic and the change of online shopping.

$$lnT_{it}^{W_t=0} = \iota_i + \theta D_t^{W_t=0} + \beta_1 lnO_i D_t^{W_t=0} + \beta_2 lnS_i D_t^{W_t=0} + \epsilon_{it}^{W_t=0} \tag{44}$$

Here, I expect that $\beta_1 = \beta_2 = 0$.

4.4.3 Further Differencing Out Unobserved Trends: Triple Differences in Style

Combining equation 43 with equation 44 provides another difference-in-differences in style or triple-differences in style estimate of the effect of online shopping on traffic congestion. The control group includes the week before the online shopping and the week before the offline shopping events (in different months), while the treatment group includes the two weeks after both events. The post online shopping event week in the treatment group is treated (by the online shopping event). Taking the difference of equation 43 and equation 44 can eliminate unobserved city-specific monthly trends $\iota_i D_t$. For example, online price tends to start low after Chinese New Year, and increase mildly throughout the year, reaching a peak before the following Chinese New Year, then plummeting to a low point again[58]. The triple difference in style specification is,

$$lnT_{it} = \iota_i D_t + \iota_i W_t + D_t W_t + \beta_1^{triple} lnO_i D_t W_t + \beta_2^{triple} lnS_i D_t W_t + X_i + \lambda_h + \varepsilon_{it} \tag{45}$$

β_1^{triple} in Equation 45 estimates a triple differences in style estimator. Again, if online shopping is more traffic-efficient, β_1^{triple} should be negative. Note that this is different from the classical

---

[58]See Alibaba Shopping Price Indices (aSPI) `http://topic.aliresearch.com/market/aliresearch/aspi.php`

difference-in-differences specification in the literature where the treatment and control groups contain different cross-sectional observations. The control group in this setting contains the same individual cities as in the treatment group, but at different times. The time interval between the two events is about two weeks, which could arguably insulate the impact of the first event from the second event. If that is true, the two weeks surrounding the offline event can be used as the counterfactual outcomes for the two weeks surrounding the online event.

5 Initial Evidence on the Connection Between Online Shopping and Traffic Congestion

This section provides initial evidence on the connection between online shopping and traffic congestion. First, I demonstrate that the trends of intracity and intercity traffic break around the online shopping event. Second, I show that there is a substantial change in online shopping patterns around the event. Third, I provide graphic evidence on the correlation between the change in online shopping and the change in traffic congestion.

5.1 The Traffic Congestion Trend Surrounding the Event

Figure 2 compares the intracity traffic and intercity traffic one week before and two weeks after the event. To highlight the difference of traffic congestion or traffic time in different time of a day, I divide a day into five segments: Morning off-peak (before 7 am), Morning peak (7 am-10 am), Day Off-peak (10 am-17 pm), Evening Peak (17 pm-20 pm), Evening Off-peak (20 pm-0 am). Figure 2(a) shows the intracity traffic congestion in one week before and two weeks after the Singles' Day shopping event. The dashed orange line provides the average traffic congestion index one week before the event and serves as the reference group. The solid blue line moves downward in most segments, which suggests that traffic congestion within cities eased in the first week after the event. The short-dashed green line shows average the traffic congestion index in the second week after the event, which tends to be regressive towards the week before the event. Figure 3 plots the de-trended traffic congestion index[59] in the three weeks and highlights the impact of the event on traffic congestion. These changes of traffic congestion index and travel time index in the upper panel and the bottom panel suggests that this short-term surge of online shopping reduces traffic within cities and increases traffic in the intercity roads, which provides suggestive evidence for the prediction from equations 27 and 30. Figure 2(b) shows the trend of intercity traffic congestion surrounding the event. The y-axis is the travel time index, which is obtained following two steps. First, I regress the raw intracity travel time data on hour $\times$ day of week fixed effects to obtain the

---

[59]I regress the traffic congestion index on hourly and day of week fixed effects first and then take the residual.

residuals of the regression. I then add the mean of the raw travel time in the regression sample to the residual. Using the index instead of raw data is to address the missing variable problem. In contrast to Figure 2(a), the solid blue line that represents the average travel time index one week after the event moves up substantially in all time segments in a day, and falls to the levels closer to the pre-event week in the second week after the event.

Table 3 quantifies the change of traffic within cities following the specification of equation 35. The first three columns show the result for peak hours and the last three columns show the result for off-peak hours. The coefficient of $Week_1$ dummy in column 1 shows that peak hour traffic congestion is reduced by 3.3% in the first week. The coefficient of $Week_2$ dummy shows that the traffic congestion index bounces back to the level before the event in the second week after the event. Column 4 shows the corresponding result for the off-peak hour sample. The traffic reduction effect is about one-third of that in the peak hour sample. Interestingly, there is a small increase in traffic congestion in the second week. Given the seasonal trends in traffic congestion, it is difficult to interpret this slight bounce in traffic congestion. Columns 2 and 5 add the interaction terms of week dummies with the log online shopping and the log online selling index. I find that cities with a higher online shopping index experienced a larger reduction in traffic while cities with a higher online selling index experienced a higher increase in traffic, with stronger effects in the peak hour sample. Columns 3 and 6 further control for other city characteristics that might affect the change in traffic congestion due to the event by adding the interactions of the week dummies with these characteristics. Log GDP per capita is used to control for income. Log number of mobile users and internet users are used to control for the number of online consumers. Again the change in traffic is negatively correlated with the online shopping index while positively correlated with the online selling index. Similar results are found in off-peak hours, although the estimate for the interaction term of $Week_1$ dummy and the log online shopping index reduces traffic and is less precise. These results provide strong evidence indicating that cities engaging in higher intensity of online shopping experience more reduction in traffic congestion. Appendix Table D.2 shows the regression results of the changes in intercity traffic[60] following equation 36.

5.2   The Trend of Online Shopping Surrounding the Event

Next, I turn to measure the change of online shopping in the weeks before and after the event using the Baidu Index as a proxy. The search index is consistent with the online shopping index

---

[60]The intercity comparison uses data in the year 2018. The measure is travel time in minutes. The intracity comparison uses data in the year 2016. The measure is traffic congestion index. It would be ideal to know the intercity travel time in the year 2016, but I started to collect real-time intercity travel time data this year. I did not use intracity traffic congestion data in the years 2017 and 2018, because offline retailers also participated in the Singles' Day Shopping Event.

provided by the online shopping platform. Figure 10 shows that the logarithm of Baidu Index is positively correlated with the logarithm of online shopping index in an expanded sample consisting of 277 cities, both before and during the shopping event. Table 4 quantifies the linear relationship between the two indices, controlling for city characteristics. The table presents the results in both the regular sample where I have the traffic congestion data and the expanded sample where I have the online shopping index and the Baidu Index but not the traffic congestion data. In the regular sample, cities with a 100 percent higher online shopping index have, on average, a 63 percent higher value of the Baidu Index in normal days (i.e., non-sales event days). This correlation is reduced by two-thirds after controlling for income and online consumers in Column 2. The Baidu Index increased by about 1.6 times ($exp(0.915) - 1$) during the event. Column 3 shows that the cross-sectional correlation between online shopping index and Baidu Index holds up during the event. The interaction of the event dummy and the log online shopping index is positive but not significant, indicating that cities with higher online shopping index have a higher increase in online shopping. Columns 4-6 show the same results in the expanded sample as Columns 1-3. The pattern remains and the interaction of the event dummy and the log online shopping index is larger and statistically significant as the sample size increases, which suggests that cities that are more adapted to online shopping spent even more during the sale. This is important for interpreting the results presented earlier in Section 5.1 and that I will show in Section 6.4, where I use the interaction of the event dummy and the online shopping index as the key regressors. This rules out the possibility that online shopping grows less in cities with higher online shopping index due to mean reversion. Thus, this interaction term contains the variation of differential growth of online shopping due to the event across cities.

5.3 The Connection Between the Change in Online Shopping Activity and the Change in Traffic Congestion

Does the change in online shopping activity correlate with the change in traffic congestion? Figure 4 presents the scatter plot of the change of the logarithm of peak hour traffic congestion against the growth rate of the Baidu Index in one week before and one week after the event. Most cities experienced a drop in traffic congestion after the event. The magnitude of the reduction effect is larger for cities with a higher increase in online shopping as indicated by the dashed line with a modest negative slope. Given that the online shopping event is the only significant event in the narrow two-week window, this provides strong suggestive evidence on the connection between the increase in online shopping and the reduction in traffic congestion. The next section further investigates the potential causal link.

34

6 Regression Estimates of the Effect of Online Shopping on Traffic Congestion

6.1 OLS Estimates

Table 5 reports the OLS estimates for the effect of online shopping on traffic congestion. The dependent variable is log traffic congestion, and the key regressor of interest is the Baidu Index in time t divided by its value in $t = 0$. The OLS results indicate a negative association between online shopping and traffic congestion. Column 1 follows equation 32 and estimates a regression model without a common trend in traffic congestion. The following columns estimate equation 38, which accounts for common trends. The sample in Column 2 contains all hours, while Columns 3 and 4 look at peak hours and off-peak hours separately. The result from peak hours is much stronger than the effect estimated from off-peak hours. As the Singles' Day shopping event in 2016 falls on a Friday in the first week and some consumers may adjust their travel plans in order to have enough time to shop on the internet[61], Column 5 excludes Fridays from both pre-event and post-event weeks. The result remains. Finally, the last column excludes both Fridays and off-peak hours, which is the preferred sample specification and the sample will be used in the IV estimation. It suggests that a 10% increase in online shopping reduces traffic congestion by 0.13%. Since the traffic congestion index is likely to correlate within cities in the time dimension, I cluster the standard errors by cities in all columns[62].

Given the granularity of the data in time, I stratify the data by hours and plot the β in equation 38 in Figure 5. The traffic reduction effect is most significant from 9 am to 10 am and around 7 pm. This is in line with the prediction of the theoretical model. Because the ratio of traffic density to free-flow density $\frac{n}{n_m}$ in peak hours is much higher than off-peak hours, the size of the effect in peak hours should be larger. Intuitively, traffic congestion is more likely to happen when the sum of different types of trips, such as shopping trips, commuting trips, leisure and other trips, exceeds a threshold where the road's maximum traffic capacity is reached[63]. Commuting trips have consumed most of the road capacity and left the roads in a congested or semi-congested situation in peak hours. Therefore the impact of the reduction in offline shopping trips on traffic congestion

---

[61]The company used live stream to engage consumers during the event. Consumers may go home early to participate in a series of interactive sales.

[62]Estimating the first-differences specification

$$\Delta ln T_i = \beta \tilde{B}_{it} + \tau_i + \varepsilon_i$$

gives very similar point estimates.

[63]See Braithwaite (2017) for Figure 4.3, which shows personal trips by start time and purpose in weekdays in England in the year 2011. Shopping trips are a non-negligible component of traffic in peak hours.

is much more notable.

6.2 The Instrumental Variable Estimates

Table 6 reports the results from estimating equations 40 and 41. The dependent variable is the change of the logarithm of the weekly average traffic congestion index. Specifically, I first take the mean of congestion index in peak hours from Monday to Thursday (excluding Friday to avoid the event day) for each city and week. I take the logarithm and then take the difference across weeks. As shown in Column 1 in the upper panel, the waived postage fee is a strong predictor of the increase of online shopping conditional on log market potential. I further add second-order and third-order polynomial terms of log market potential in Columns 2 and 3 to strip out any variation in the IV that is related to remoteness, trade quantity and price index. In Column 4, I add the control for average income measured by GDP per capita, and the coefficient of interest barely changes. Column 5 further controls for the number of mobile users and internet users. The combination of both variables controls for the number of online consumers. The waived postage fee remains highly significant through all specifications. The F-Statistics in Columns 1 and 2 indicate that the first-stage impact of waived postage fee is very powerful, despite that it drops when adding more potentially irrelevant controls in Columns 3-5. The second panel of Table 6 shows the reduced-form result of the effect of waived postage fee on traffic congestion. Waiving postage fee has a significant and consistent effect on the reduction of traffic congestion. Taking literally, these results imply that the interaction of the reduced postage fee and the online shopping event resulted in a surge in online shopping and a reduction in traffic congestion.

Table 7 contains the instrumental variable estimates of the effect of online shopping on traffic congestion with three different constructs of market access. In the first panel, the market access is the inverse distance weighted city population, so the reported β^{IV} is simply the ratio of the reduced-form estimates in Table 6. The estimates suggest that a 10% increase in online shopping reduces traffic congestion by 1.4%-1.7% in peak hours, which is an elasticity of -0.14 to -0.17. Further, these estimates are largely insensitive to polynomial terms of different orders of market access and including control variables on income and the number of online consumers. The results are also robust to different measures of market access. The second panel uses the number of both Tmall and Taobao online stores by cities listed in the online shopping platform of Alibaba and the last panel uses the number of only Tmall shops as the market access. The results support the robustness of the IV estimates in the first panel. Although not reported here, the elasticity of traffic congestion to online consumption for all hours is -0.094 to -0.113.

Note that the IV estimate is much larger than the preferred OLS estimates in Table 5. As the study

by Løken et al. (2012) shows, the difference between the OLS estimates and IV estimates can be decomposed into two parts: the difference in the *marginal effects* between OLS and IV estimates, and the difference in the *weights* between OLS and IV estimates. Reasons that could explain the differences in the marginal effects have been discussed earlier in Section 4.2, including measurement error, reverse causality and omitted variable bias. Here, I explore the possible difference between the OLS and IV estimates arising from the difference in the regression weights following the method proposed in Løken et al. (2012). The regression weights can be calculated as,

$$w_{gi}^{OLS} = \frac{Cov(d_{gi}, \tilde{B}_i)}{Var(\tilde{B}_i)} \tag{46}$$

$$w_{gi}^{IV} = \frac{Cov(d_{gi}, P_i)}{Cov(\tilde{B}_i, P_i)} \tag{47}$$

where $\tilde{B}_i$ is the Baidu Index growth rate, which is the endogenous variable in regression 42. P_i is the instrumental variable: the waived postage fee. d_{gi} are dummy variables constructed as $d_g = 1\{\tilde{B}_i >= b\}$. b are evenly distributed cutoffs with an interval of 0.1 that divides the range of $\tilde{B}_i$ into 20 groups, with g indicating each group. w_{gi}^{OLS} and w_{gi}^{IV} gives the weights of OLS and IV estimates in group g, respectively. As shown in Løken et al. (2012), OLS estimates give more weights to the marginal effects close to the sample median of the regressor, while IV estimates assign more weights to the marginal effects for the changes of the endogenous variable that are most affected by the IV. Given the heterogeneity in β, the IV weights lead to different estimates even if the marginal effects of OLS and IV are the same. Figure 6 shows the distribution of the weights IV and OLS estimates, respectively. The Figure reveals that the IV estimate assigns more weights to the marginal effects in the right tail of the distribution of the growth rate of Baidu Index relative to the OLS estimate. A following up question is what are the characteristics of the cities in the right tail. Appendix D.1 provides some suggestive evidence showing that the IV estimates are identified from the cities with higher market potential, income, and number of online consumers.

6.3 Heterogeneity

My theoretical model predicts that the elasticity ϵ_j include three components: congestion level $\frac{n_j}{n_{mj}}$, the importance of online shopping traffic $\frac{\psi_j \pi_o}{\delta_j \pi_o + \zeta_j \pi_f}$, and the traffic saving factor. All three parts could vary with cities, which results in heterogeneity in β. Among the three sources of heterogeneity, the comparison of the effect between peak hours and off-peak hours has shown the heterogeneity due to the first component $\frac{n_j}{n_{mj}}$. For the second component, I do not have data on the share of online consumption (π_o) or the share of shopping-related traffic (ψ), or the share of shopping made through private vehicles (ζ) by cities. This section focuses on the investigation of

the heterogeneous effect caused by the third component: the traffic-saving factor.

The theory predicts that the reduction of traffic congestion is caused by higher traffic efficiency in online shopping relative to offline shopping. If that is correct, the more products delivery vans can carry, the more traffic-efficient is online shopping. The number of products that a van of standard size can carry depends on the size of goods; therefore, online delivery of bulk goods may have lowered traffic efficiency relative to smaller pieces of goods such as apparel. This implies that δ varies across product categories. I do not have data on the volume of purchase by categories in cities; however, I extracted the number of online shops by categories from the website of the online shopping platform. In the international trade literature, countries are found to be the net sellers of the products that they demand the most, which is known as the home market effect. Analogously, if such effect also exists in the trade across cites, then the number of online store of certain category could indicate the domestic demand for the products in that category in the registration city (Costinot et al., 2016; Coşar et al., 2018). Assuming that the number of online shops in certain categories registered in a city provides a proxy for the consumption of the category[64], I might obtain some suggestive evidence of how differential traffic-saving factors across product categories causing heterogeneous congestion relief effect. For that reason, I add interaction terms of the change in online shopping with the number of online shops in a certain category in cities to equation 42. Formally, I estimate

$$\Delta ln T_i = \alpha^{IV} + \beta^{IV} \hat{\bar{B}}_i + \beta^{IV_{Int}} \hat{\bar{B}}_i H_{ik} + f(M_i) + X_i \theta^{IV} + \varepsilon_i$$

, where H_{ik} is the number of online shops in category k in city i.

I adopt two strategies to estimate the endogenous interaction term $\hat{\bar{B}}_i H_{ik}$. The first IV strategy interacts the waived postage fee with the number of online shops in categories to instrument the endogenous interaction term. The second IV strategy first predicts the growth rate of the Baidu Index after regressing it on the waived postage fee and controls, and then use the interaction of the predicted value with the number of online shops in category k as the IV for the endogenous interaction term. Table 8 shows the result for the heterogeneous effects estimated in both ways. The upper panel presents the results estimated with the first strategy, and the lower panel reports the second. There is little difference between the results of the two strategies. Cities with a higher number of online shopping in the categories of furniture, appliance and home products tend to benefit less from the increase in online shopping. This is consistent with the prediction from the theory: the products in these categories are bulky and have lower traffic saving potential ($|\delta - \frac{\zeta_j}{|\lambda|}|$ is lower).

---

[64]Here, the location choice of retailers indicates local demand for certain products.

6.4 Event Study Estimates

Table 9 shows the results for the event study estimation discussed in Section 4.4.3. The dependent variable is log traffic congestion index. As explained in Section 5.2, the interaction of online shopping index with period dummy captures differential growth of online shopping across cities. I control for city average income and number of online consumers by including the interactions of these variables with the period dummy. Standard errors are clustered at city level to address the serial correlation of traffic congestion in time within cities. The first panel compares the traffic one week *after* the online shopping event with the traffic one week *after* the offline shopping event following equation 43. Consistent with the earlier results, the cities with higher online shopping activities experienced a larger reduction in traffic congestion. A 100% increase in online shopping index reduces traffic congestion index by 1.9%, on average. The effect is stronger over peak hours (3.3%) and weaker (1.5%) during off-peak hours. Note that the magnitude is not comparable with the earlier results as the measure of online shopping is different. Interestingly, the online selling index appears to have a positive effect on traffic. I interpret this result as reflecting the increased traffic due to online delivery in the cities that are more actively engaged in online selling activities.

The second panel compares the traffic one week *before* the online shopping event with the traffic one week *before* the offline shopping event following the specification in equation 44. When the event is "turned off", the correlation between the change in traffic and the change in online shopping disappeared. This assures that the connection between traffic and online shopping event is not just a superficial correlation. The third panel estimates equation 45 which takes the difference of the first two panels to eliminate potential monthly trends. The main result reduces slightly, but remains significant. A 100% increase in online shopping index reduces traffic congestion index by 1.4%, with a larger effect for peak hours (2.5%) and a smaller effect for off-peak hours (1.1%). The online selling index continues to correlate with traffic positively, but the effect loses its statistical significance.

6.5 The Effect on Air Pollution

Another question of interest is whether online shopping leads to the reduction of traffic-related air pollution, which is a proposed mechanism in Gendron-Carrier et al. (2018). Table 10 shows the results of estimating equation 38 using air quality index and six other types of air pollutants as dependent variables. The first takeaway is that there is a common trend of increasing air pollution in the week after the event for the seven indicators except for Ozone. The point estimates for the effect of online shopping on the change of NO2 and CO suggest that online shopping reduces traffic-related air pollution; however, the effects are not statistically significant. Online shopping

appears to have caused little changes in the other air pollution indicators. One possible explanation is that the surge of online shopping drives up products production, which could contribute to the common trend of increasing air pollution. It is also possible that delivery vans use diesel and produce higher per-vehicle air pollution relative to private vehicles, which may have offset the reduction effect of online shopping on traffic congestion.

7 Welfare Analysis

If we divide the purpose of travels into commuting, shopping and others, then the reduced travel demand for shopping will benefit the other two types of travels. I focus on the congestion relief benefit for commuters in peak hours, because I do not have data on the number of travels for specific purposes and commuting travels are the majority.

Figure 7 presents a classic static economic model of traffic congestion. The downward-sloping blue curve is the travel demand function. The upward-sloping green curve is the average cost of trips. The per-trip cost increases with the number of vehicles on roads due to the externalities of traffic congestion. Note this curve is flat when traffic density is sufficiently low. The constant marginal cost of each trip includes costs such as time cost, fuel cost, or fare of a bus trip. The slope of the curve becomes upward when traffic density is higher than free-flow density, when an additional vehicle causes delay for existing vehicles. Given the focus here is traffic congestion in peak hours, I assume that the average costs of trips increase with the number of trips. At the original equilibrium point (q_1, p_1), the total welfare of consumers from these travels is $A + B$. The welfare loss due to negative externality from traffic congestion is $C + D + E$[65]. When e-commerce takes away part of the shopping travel demand, the demand curve shifts inward to the orange line. A new consumer equilibrium will be reached at point (q_2, p_2). The total welfare of consumers from these travels is then $A + C$, and the welfare loss reduces to E. The welfare gain is then $C - B(=(A + C) - (A + B))$. Note that B actually represents the welfare of consumers who switched from offline shopping to online shopping. Assuming offline shopping and online shopping provides the same level of utility, B is negligible. The welfare gain is then just C. If we know the decrease of time cost of travel $p_1 - p_2$, and the number of trips q' at the new equilibrium, then we can approximate part C using the shaded square in the graph (the region of p_1, a, b, p_2). The triangle part a, b, c represents the welfare gain due to the adjustment of travel demand in the long run. Given that I cannot clearly identify the long-term effects for reasons listed in Section 8, I will only focus on the short-term effect here.

---

[65]This obtained by integrating the marginal negative externality along the upward-sloping green curve up to the point of the market equilibrium.

I estimate the welfare improvement only for Beijing. Apart from data constraint, congestion relief in Beijing has been extensively studied and these prior estimates provide a benchmark to compare the effect of e-commerce with other potential policies such as congestion charge on the demand side and providing new subways on the supply side. According to Gu et al. (2019), the average commuting time is 56 minutes one-way in Beijing in the year 2016. The average peak hour traffic congestion index before the event is 2. This implies that the free-flow travel time is 28 minutes for a typical commuting trip. If the congestion index goes down by 1.4%, then the actual travel time will be 55.2 minutes, and the travel time saving is 0.78 minutes for each way. Therefore a 10% increase in e-commerce saves 1.57 minutes per workday. The value of commuting time can be derived from the annual wage and working hours, discounted by a factor of 0.5 (Anderson, 2014), which is 0.77 *yuan*/minute. Beijing has 5.7 million people commuting by car and 5.2 million people commuting by bus in a workday. Assuming cars and buses are affected by traffic in the same way, and there are 250 workdays in a year, the estimated welfare gain *C* will be 1.65 billion *yuan* for 239 million US dollars for peak hours. This is about a third of the size of the gain estimated from the supply of new subway lines in Gu et al. (2019). However, the cost of this "policy" appears to be much smaller than the heavy investment in infrastructure, because this is achieved by improving the traffic efficiency in shopping goods within cities and the change of shopping behavior. There is potentially rising cost of traffic congestion in the intercity roads as shown in this paper; however, its aggregated value is likely to be small given the share of population traveling on the highway for commuting is much smaller relative to commuting on intracity roads. There are obviously many other gains and losses in a broader sense, such as benefits from the convenience of online shopping, losses due to the structural change of the economy. However, these factors are not directly related to traffic congestion, and thus they are out beyond the scope of this paper.

8   Discussion on the Long Run Effect

Above results show that traffic congestion was reduced in the short-term event. Does the growing popularity of online shopping improve traffic congestion in a longer term? Through the lens of the theoretical model, the empirical results suggest that the traffic-saving factor is smaller than zero. The interpretation of the result is that the vehicle-saving ratio is sufficiently low, and the amount of offline shopping that consumers are willing to substitute with online shopping is sufficiently large. In the long-run, the vehicle-saving ratio is likely to become even lower. As online shopping is increasingly convenient, the amount of offline shopping that consumers are willing to give can be even larger. Therefore online shopping is likely to continue to reduce the overall shopping vehicle demand. However, adaptation may reserve this result. Similar to the proposition of the fundamental law of road congestion by Duranton and Turner (2011), the reduction in shopping-related vehicle

demand might be offset by vehicle demand for other purposes, such as commuting and leisure trips. Without knowing the long-term travel demand, I cannot predict the general equilibrium effect of online shopping. It is possible to answer the question empirically using longer-term changes in online shopping and traffic congestion. Unfortunately, I currently do not have such data. For the change in online shopping, the Baidu Index does not appear to be measuring the long-term change of online shopping. The official statistics indicate that the overall online sale increased by 32.2% between the year 2016 and 2017[66]; however, there is little increase in the Baidu Index. Nevertheless, Gu et al. (2019) provides some positive perspectives on the long-term effect. They find a persistent congestion relief effect of new subway lines on traffic congestion over the timeframe of one year. This suggests that the reduction of traffic achieved through diverting peak-hour road traffic to other transportation modes can be larger than the latent travel demand. E-commerce could at least be a complementary soft policy that reduces traffic congestion along with many other congestion relief policies such as congestion charge and provision of the fast transit systems in dense urban areas.

9 Conclusions

The paper has aimed to make three contributions to our understanding of the congestion relief effects of e-commerce in the context of the Singles' Day Shopping Event. First, I derived the elasticity of traffic congestion index to the quantity in online shopping. The elasticity includes three components: traffic density, the importance of e-commerce, and a traffic-saving factor, which is per-unit online good traffic-saving. The condition for online shopping to reduce traffic is the traffic-saving factor being negative. Quantifying the traffic-saving factor depends on knowing the substitution between online and offline channels, which relies on assumptions of channel choices on the demand side.

My second contribution is to develop such a model that can predict the online-offline substitution of quantities. Portraying e-commerce as trade across cities, I develop a trade model with heterogeneous consumers and two shopping channels. Inspired by the mechanism that the lowest price wins the market introduced by the work of Eaton and Kortum (2002), I assume that consumers have Fréchet distributed matching quality with varieties, and purchase from the channel that gives the higher matching quality. The model thus inherits the beauty in the price distribution properties from Eaton and Kortum (2002) model: the distribution of quantity consumed in a specific channel is independent of the channel. The channel specific quantity consumed can be expressed as the mean quantity consumed and the share of consumers that choose that channel. The effect of the price change on quantity in each channel can then be decomposed into a mean effect and a share effect.

---

[66]See http://www.ec.com.cn/article/dssz/scyx/201801/24827_1.html

Particularly, the online-offline substitution is found to be a concise function of the elasticity of online consumption quantity to the relative price of online to offline channel, and the elasticity of substitution between varieties.

My third contribution is exploiting the traffic congestion reduction induced by the event to provide the first available empirical evidence on the congestion-relief effect from e-commerce. The evidence suggests that the reduction in the traffic congestion index surrounding the event is related to the increase in online shopping. The temporarily waived postage fee during the event day provides exogenous incentives for consumers to switch to online shopping. Using the waived postage fee as an instrument, I estimate that a 10% increase in online shopping reduces traffic congestion by about 1.4%, which is about a -0.14 elasticity. While the anticipation of the event may encourage consumers to hold up consumption until the event day, and thus introduces bias, I find the congestion relief effects of online shopping remain significant in a difference-in-differences specification, where the counterfactual traffic congestion level is the week after another a large scale shopping event. The effect is stronger in peak hours when traffic density is high, and in cities that engage in more online sales of bulk product, which confirms the embedded heterogeneity of the elasticity as predicted by the theoretical model. Welfare calculations suggest that the reduced shopping vehicle demand leads to a welfare gain of about 239 million US Dollars for peak hours, which is about a third of the size of the welfare gains estimated from the supply of new subway lines in a city.

Admittedly, at least three limitations are worth noting. First, I use the trend in the searches of the online shopping platform on the internet monitored by Baidu Index as a proxy of the growth rate of online shopping quantity. The growth rate of the Baidu Index may be systematically higher or lower than the actual growth rate of online shopping quantity and leads to bias in the estimate of the elasticity. Obtaining confidential online shopping data by cities recorded by the online shopping platform can improve the estimate. Second, because the event creates a spike in online shopping, the empirical results may not reveal the effects of the marginal changes in the adaptation to e-commerce. Third, it is difficult to generalize the effect from the day-long shopping event to effects over longer periods of time, without knowing the travel demand elasticity over a longer horizon. Online shopping can reduce shopping vehicle travel demand but may not necessarily reduce traffic congestion due to potential long-run adaptations as predicted by the fundamental law of traffic congestion. Exploring the long-term congestion reduction effects of e-commerce may be an area of future research.

Table 1: Parameter values for quantitative analysis

Parameter	Variables	Mean	Source	Year
Statistics				
Sample mean of Online shopping growth rate	B	1.6	Baidu Index data	2016
Mean change in price shock	k	0.25	CCX Credit Technology Online Report	2016
Sample mean of traffic congestion index	T	1.65	Traffic congestion data	2016
The ratio of traffic density to optimal traffic density	$\frac{n}{n_m}$	2.3	Traffic congestion data	2016
Share of online shopping	π_o	0.126	National Bureau of Statistics of China	2016
Share of shopping made through private cars	ζ	0.86	US National Household Travel Survey (NHTS)	2017
Share of shopping vehicles to the overall number of vehicles on the road	ψ	0.31	US National Household Travel Survey (NHTS)	2017
Parameters				
Elasticity of substitution	σ	4	Redding and Sturm (2008)	
Per unit good vehicle-saving ratio	δ	0.067	BBC News	
The ratio of reference speed to free-flow speed	u	5	Notley et al. (2009)	

Table 2: Summary statistics for traffic congestion and air pollution by peak hours

| | Peak hours | | Non-peak hours | |
	Before	After	Before	After
Traffic congestion index	1.74	1.67	1.31	1.29
	(0.33)	(0.25)	(0.22)	(0.19)
NO2	42.86	56.44	38.33	52.40
	(20.88)	(26.84)	(21.47)	(27.45)
CO	1.05	1.44	1.00	1.40
	(0.62)	(0.90)	(0.62)	(0.91)
AQI	73.26	105.90	73.58	109.42
	(51.37)	(66.74)	(54.21)	(70.45)
O3	36.11	37.89	40.09	40.40
	(25.31)	(35.91)	(26.67)	(36.04)
PM10	82.85	125.93	83.67	129.36
	(70.85)	(92.78)	(77.42)	(92.47)
PM2.5	46.50	74.09	47.47	77.43
	(36.84)	(55.79)	(38.61)	(58.97)
SO2	20.87	26.30	20.36	26.89
	(21.62)	(28.24)	(21.43)	(29.35)
Observations	2820	2820	8207	8460

Note: The observation is the city-hour. Parentheses contain standard deviations.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table 3: Estimates of the changes in intracity travel time before and after the event

	(1)	(2)	(3)	(4)	(5)	(6)
Week 1	-0.033***	-0.117***	-0.047	-0.011***	-0.048***	-0.017
	(0.003)	(0.029)	(0.055)	(0.002)	(0.015)	(0.043)
Week 2	-0.005	-0.038	0.147	0.008***	0.009	0.113
	(0.004)	(0.050)	(0.109)	(0.003)	(0.027)	(0.083)
Week 1 × Ln online shopping index		-0.016**	-0.009		-0.009**	-0.009
		(0.007)	(0.006)		(0.004)	(0.005)
Week 2 × Ln online shopping index		-0.004	0.002		-0.002	-0.000
		(0.013)	(0.013)		(0.006)	(0.008)
Week 1 × Ln online selling index		0.045***	0.036**		0.021**	0.020**
		(0.016)	(0.014)		(0.009)	(0.009)
Week 2 × Ln online selling index		0.016	0.016		0.001	0.004
		(0.028)	(0.028)		(0.015)	(0.017)
Week 1 × Log GDP per capita			0.004			0.002
			(0.005)			(0.003)
Week 2 × Log GDP per capita			-0.008			-0.007
			(0.009)			(0.008)
Week 1 × Log mobile users			-0.026***			-0.020***
			(0.007)			(0.006)
Week 2 × Log mobile users			-0.028*			-0.011
			(0.016)			(0.010)
Week 1 × Log internet users			0.013**			0.016***
			(0.006)			(0.006)
Week 2 × Log internet users			0.016			0.008
			(0.016)			(0.010)
R^2	0.614	0.612	0.614	0.842	0.841	0.841
N	11287	10807	10567	33278	31870	31159

Note: The dependent variable is ln congestion index. The omitted group is the week before the Singles' Day shopping event. Columns 1-3 show results for peak hours, and columns 4-6 show results for off-peak hours. City fixed effects, day-of-week fixed effects, and hour fixed effects are included. Standard errors are clustered at the city level.

* p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

Table 4: The relationship between web searches and the online shopping index

	Regular Sample			Expanded Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
Log online shopping index	0.632***	0.198***	0.167**	0.707***	0.285***	0.243***
	(0.057)	(0.056)	(0.064)	(0.032)	(0.032)	(0.037)
Online event dummy	0.915***	0.915***	0.829***	0.803***	0.802***	0.760***
	(0.094)	(0.055)	(0.091)	(0.050)	(0.029)	(0.028)
Online event dummy $\times$ Log online shopping index			0.063			0.084**
			(0.065)			(0.038)
Income		Yes	Yes		Yes	Yes
Online consumsers		Yes	Yes		Yes	Yes
R^2	0.575	0.857	0.857	0.646	0.890	0.892
N	184	182	182	552	538	538

Note: The dependent variable is log Baidu index. The key regressor of interest is log online shopping index, online event dummy and their interaction. Income control includes log GDP per capita. Internet controls include log number of mobile users, and log number of households with internet connection. Robust standard errors are applied.

* p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

Table 5: Ordinary least square estimates of the effect of online shopping on traffic congestion

	(1) No common trend	(2) Base line	(3) Peak hours	(4) Non-peak hours	(5) Excluding Friday	(6) Excluding Friday Peak hours
$\frac{B_{it}}{B_{i0}}$	-0.011***	-0.009**	-0.012*	-0.007*	-0.008*	-0.013*
	(0.001)	(0.004)	(0.007)	(0.004)	(0.004)	(0.007)
Week 1		-0.004	-0.021**	0.001	-0.006	-0.018
		(0.007)	(0.010)	(0.006)	(0.007)	(0.012)
R^2	0.855	0.855	0.647	0.852	0.857	0.643
N	22307	22307	5640	16667	17795	4512

Note: The dependent variable is log traffic congestion. The key regressor of interest is the Baidu Index in time *t* divided by its initial level. Column 1 compares the change of traffic congestion in one week before and one week after the event without common time trend. Column 2 allows for common time trend. Columns 3 and 4 show results for peak hours and off-peak hours, respectively. Column 5 excludes Fridays. Column 6 shows peak hour results with samples excluding Friday. City, hour and day-of-week fixed effects are included. Standard errors are clustered at the city level.

* p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

Table 6: Estimates of the impact of the waived postage fee on changes in online shopping and traffic congestion

	(1)	(2)	(3)	(4)	(5)
Online shopping changes					
Log avg postage fee	1.896***	1.838***	1.807***	1.808***	1.524***
	(0.486)	(0.468)	(0.472)	(0.470)	(0.395)
Traffic congestion changes					
Log avg postage fee	-0.265***	-0.286***	-0.292***	-0.285***	-0.259***
	(0.077)	(0.079)	(0.080)	(0.076)	(0.075)
F-Statistic	29.51	20.19	15.43	14.3	13.72
Market potential	Yes				
Market potential square		Yes			
Market potential cube			Yes	Yes	Yes
Income				Yes	Yes
Online consumers					Yes
N	91	91	91	90	90

Note: The upper panel presents the reduced-form regression results for the changes in online shopping. The dependent variable is the growth rate of the Baidu index, and the key regressor of interest is log average postage fee. The lower panel reports the reduced-form regression results for the changes in traffic congestion index. The dependent variable is the change in log traffic congestion in one week before the event and one week after the event. For each week, I first take the mean of congestion index in peak hours from Monday to Thursday (excluding Friday to avoid the event day) for each city and then take the logarithm. The key regressor of interest is log average postage. Income control includes log GDP per capita. Online consumers controls include log number of mobile users and log number of internet users. Robust standard errors are applied.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table 7: The instrumental variable estimates of the effect of increased online shopping on traffic congestion

	(1)	(2)	(3)	(4)	(5)
Population					
Baidu Index growth rate	-0.140***	-0.156***	-0.162**	-0.157**	-0.170**
	(0.053)	(0.059)	(0.063)	(0.064)	(0.074)
Tmall plus Taobao					
Baidu Index growth rate	-0.103***	-0.118***	-0.123***	-0.123***	-0.143**
	(0.037)	(0.038)	(0.041)	(0.042)	(0.067)
Tmall					
Baidu Index growth rate	-0.096***	-0.097***	-0.096***	-0.095***	-0.096**
	(0.035)	(0.034)	(0.032)	(0.033)	(0.040)
Market potential	Yes				
Market potential square		Yes			
Market potential cube			Yes	Yes	Yes
Income				Yes	Yes
Online consumers					Yes
N	91	91	91	90	90

Note: The first row of each panel indicates which market potential variable construction is used. Market potential in the first panel is constructed with inverse distance weighted city population; The second panel uses the overall number of online shops in cities; The last panel uses the number of Tmall shops in cities. The dependent variable is the changes in log traffic congestion index. For each week, I first take the mean of congestion index in peak hours from Monday to Thursday (excluding Friday to avoid the event day) for each city and then take the logarithm. The key regressor of interest is the growth rate of the Baidu Index. Income control includes log GDP per capita. Online consumers controls include log number of mobile users and log number of internet users. Robust standard errors are applied.

* p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

Table 8: The heterogeneous effects of online shopping on traffic congestion by product categories

	(1) Office	(2) Clothing	(3) Furniture	(4) Appliance	(5) Home	(6) Electronics	(7) Baby	(8) Cosmetics
Interacted instrument								
Baidu Index growth	-0.118**	-0.113**	-0.123**	-0.113**	-0.094***	-0.120**	-0.113**	-0.113**
rate	(0.049)	(0.052)	(0.052)	(0.047)	(0.035)	(0.052)	(0.049)	(0.050)
Interaction with	0.002	-0.001	0.006**	0.024**	0.009***	0.002	0.003	-0.003
product categories	(0.002)	(0.001)	(0.003)	(0.012)	(0.003)	(0.001)	(0.013)	(0.019)
Predicted instrument								
Baidu Index growth	-0.118**	-0.112**	-0.123**	-0.113**	-0.094***	-0.119**	-0.113**	-0.113**
rate	(0.049)	(0.053)	(0.052)	(0.047)	(0.035)	(0.052)	(0.049)	(0.050)
Interaction with	0.002	-0.001	0.006**	0.023*	0.009***	0.002	0.002	-0.004
product categories	(0.002)	(0.001)	(0.003)	(0.012)	(0.003)	(0.001)	(0.015)	(0.021)
N	90	90	90	90	90	90	90	90

Note: The coefficients are estimated using 2SLS. The dependent variable is the changes in log traffic congestion index. For each week, I first take the mean of the congestion index in peak hours from Monday to Thursday (excluding Friday to avoid the event day) for each city and then take the logarithm. The key regressors of interest are the growth rate of the Baidu Index and its interaction with the number of online sellers of the product category in the column titles. Controls variables are the same as in the last column of table 7. Income control includes log GDP per capita. Online consumers controls include log number of mobile users and log number of internet users. Controls also include third-degree polynomials of market potential. Robust standard errors are applied.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table 9: Difference-in-differences in style estimation of the effect of online shopping on traffic congestion

	All hours (1)	Peak hours (2)	Non peak hours (3)
Post week			
Period × Ln online shopping index	-0.019**	-0.033**	-0.015**
	(0.008)	(0.014)	(0.006)
Period × Ln online selling index	0.033**	0.056**	0.025**
	(0.016)	(0.027)	(0.013)
N	12672	3168	9504
Prior week			
Period × Ln online shopping index	-0.005	-0.008	-0.003
	(0.009)	(0.015)	(0.008)
Period × Ln online selling index	0.015	0.024	0.011
	(0.018)	(0.033)	(0.016)
N	12436	3168	9268
Difference			
Period × Post × Ln online shopping index	-0.014**	-0.025**	-0.011**
	(0.006)	(0.012)	(0.006)
Period × Post × Ln online selling index	0.018	0.031	0.014
	(0.013)	(0.028)	(0.011)
R^2	0.881	0.683	0.880
N	25108	6336	18772

Note: The dependent variable is log traffic congestion index. The first panel compares the traffic one week after the online shopping event with the traffic one week after the offline shopping event; The second panel compares the traffic one week before the online shopping event with the traffic one week before the offline shopping event; The third panel estimates the differences between the estimates in the first two panels. Controls variables include the interaction of period dummy with log GDP per capita, log mobile users, log internet users. The first and second panel includes period-fixed dummy and period × city fixed effects. The third panel includes the interaction of the period dummy and post-event week dummy, period × city and post-event week dummy × city fixed effects. Standard errors are clustered at the city level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table 10: Ordinary least square estimates of the effect of online shopping on air pollution

	(1) ln aqi	(2) ln co	(3) ln no2	(4) ln o3	(5) ln pm10	(6) ln pm25	(7) ln so2
$\frac{B_{it}}{B_{i0}}$	0.086	-0.035	-0.039	0.140	0.089	0.085	0.040
	(0.068)	(0.065)	(0.046)	(0.125)	(0.075)	(0.093)	(0.063)
Week 1	0.268**	0.333***	0.348***	-0.335*	0.335***	0.338**	0.156
	(0.109)	(0.106)	(0.076)	(0.191)	(0.119)	(0.152)	(0.106)
R^2	0.552	0.540	0.606	0.534	0.573	0.501	0.648
N	20084	20084	20083	20054	19975	20084	20084

Note: The dependent variables are AQI and another six types of pollutants. The key regressor of interest is the Baidu Index in time t divided by its initial level. City, hour and day-of-week fixed effects are included. Standard errors are clustered at city level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.
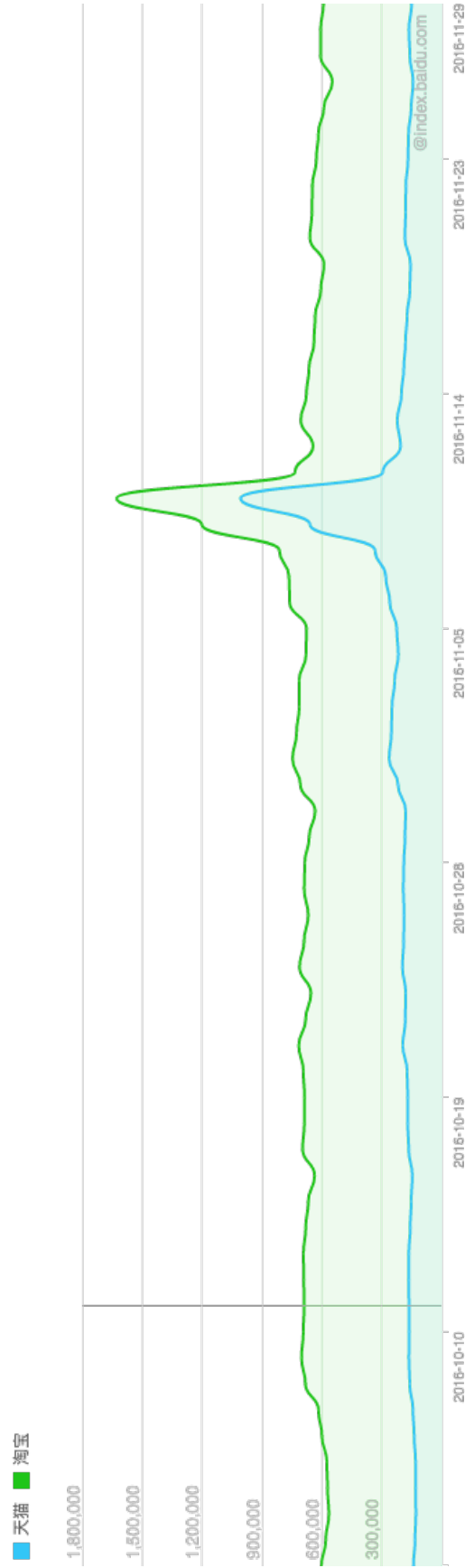
Figure 1: Baidu Index between October and November 2016

Note: The *y*-axis shows the value of Baidu Index, and the x-axis shows date. The green line on top shows the index for Taobao shops, and the blue line in the bottom shows the index for Tmall shops.
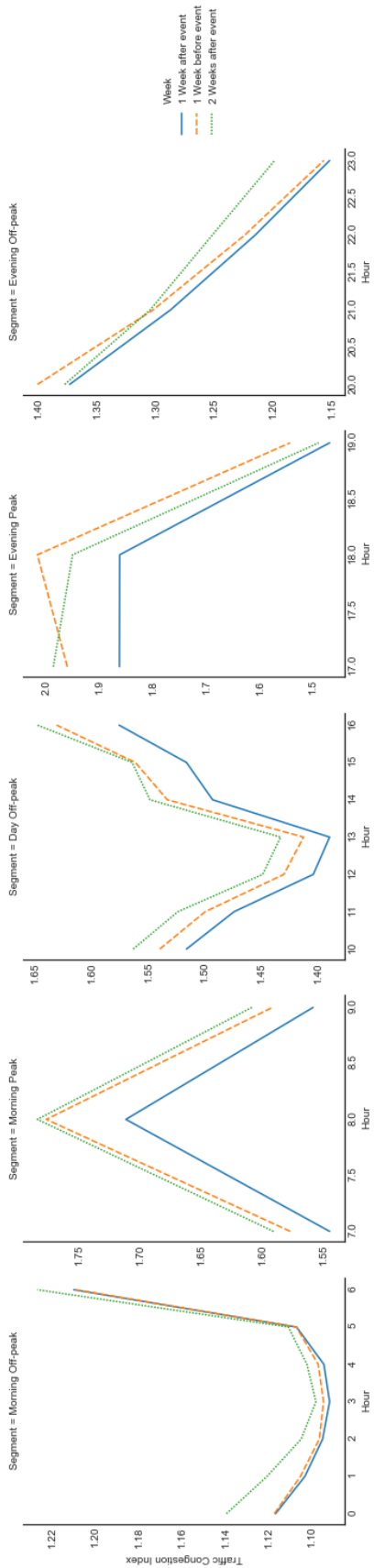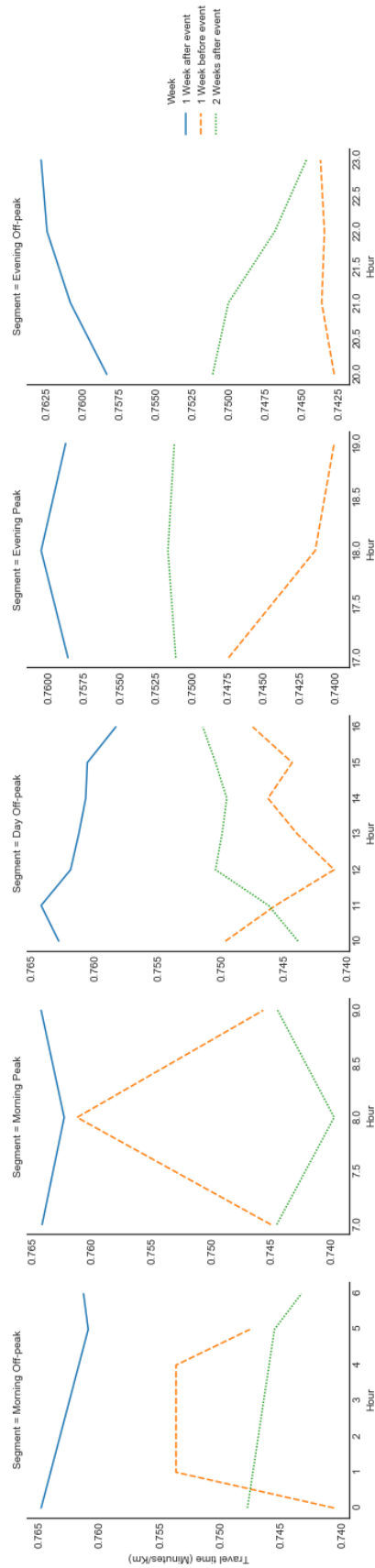
Figure 2: The changes in intracity and intercity traffic congestion surrounding the event

Note: Figure (a) shows intracity traffic congestion during weekdays in one week before and two weeks after the Singles' Day shopping event. Figure (b) shows smoothed intercity travel time in one week before and two weeks after the Singles' Day shopping event. The smoothed measure is obtained through regressing the raw traffic data on the hour × day-of-week fixed effects and adding the mean of the raw travel time in the regression sample to the residual. The graph includes observations in all weekdays in one week before and two weeks after the event. Smoothing is used to address missing observations. Travel time per km shifted upwards in the week after the event and further reduced in the following week.

54

Figure 3: The trend of traffic congestion index surrounding the event
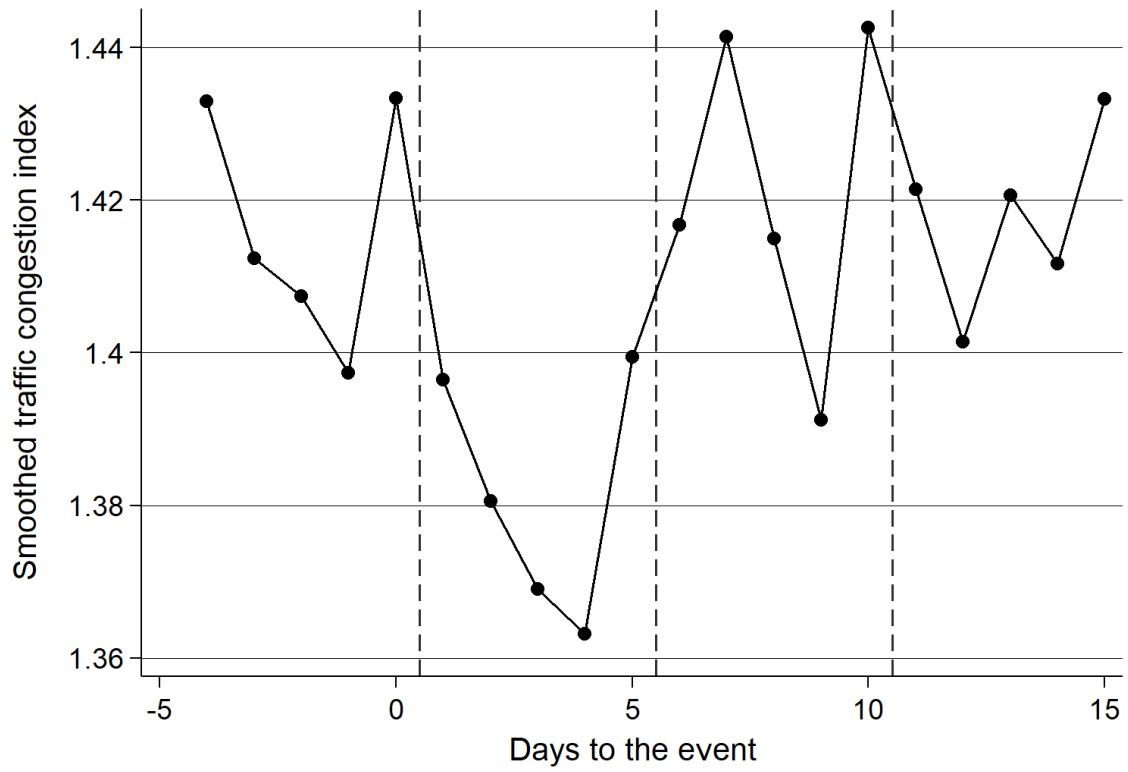
Note: Y-axis shows the residual after regressing congestion index on city, hour and weekday fixed effects. X-aixs shows the days until the event. Dashed vertical lines separate weeks.

Figure 4: Change in traffic congestion versus change in online shopping
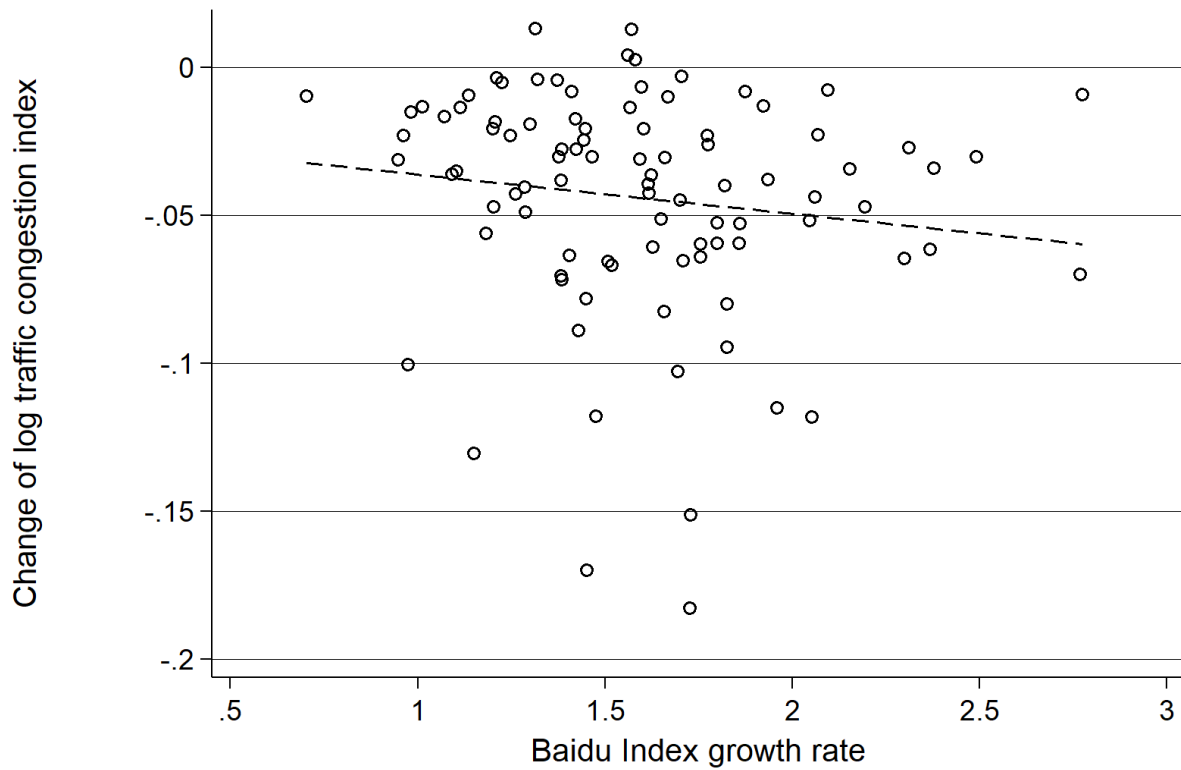
Note: The dependent variable is changes in log traffic congestion index. For each week, I first take the mean of the congestion index in peak hours from Monday to Thursday (excluding Friday to avoid the event day) for each city and then take the logarithm. The key regressor of interest is the growth rate of the Baidu Index.

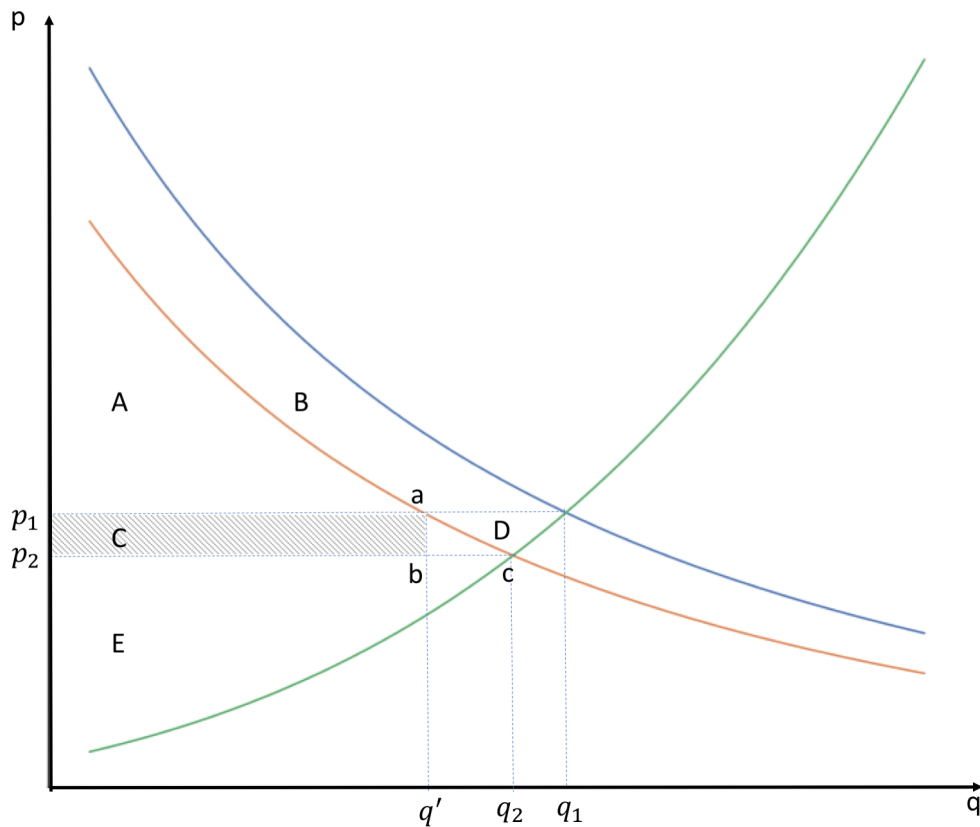Figure 5: The effects of the increase of online shopping on traffic congestion, by hour
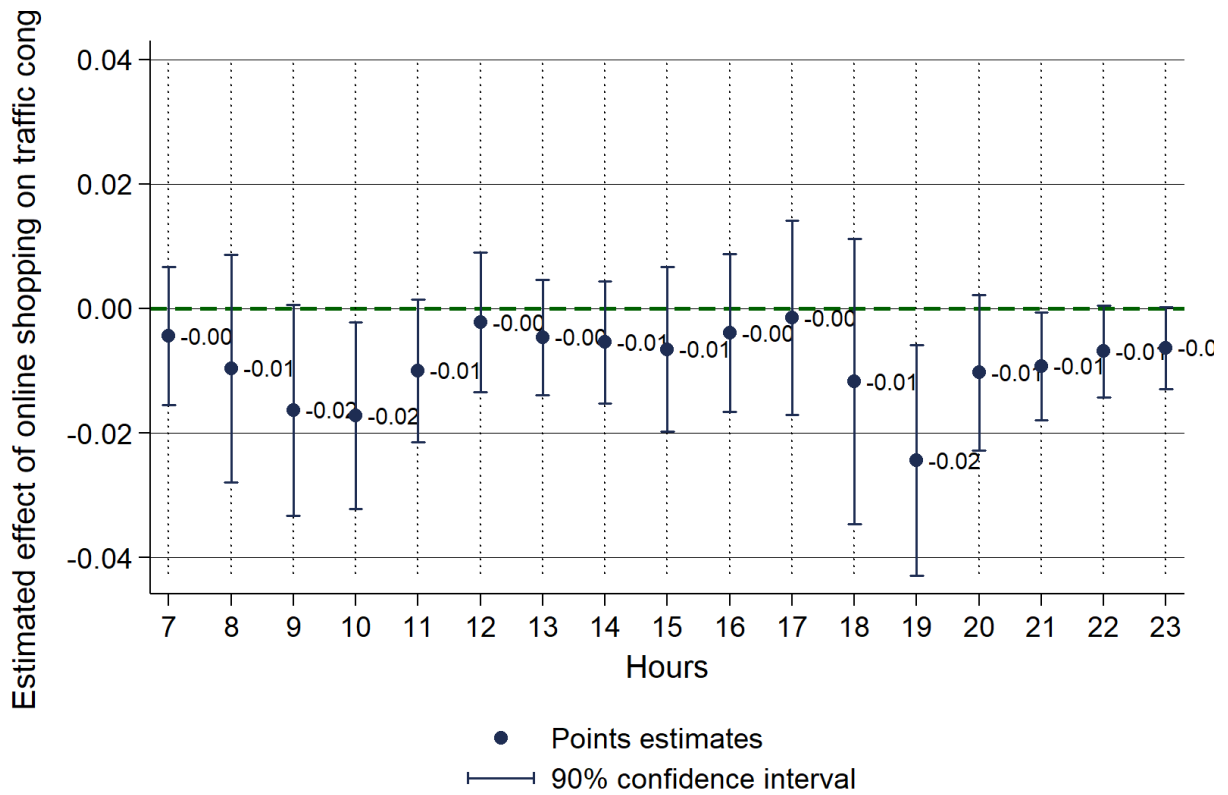
Note: This figure stratifies the data by hours and plots the coefficient β in equation 38.

Figure 6: OLS estimates and IV estimates weights
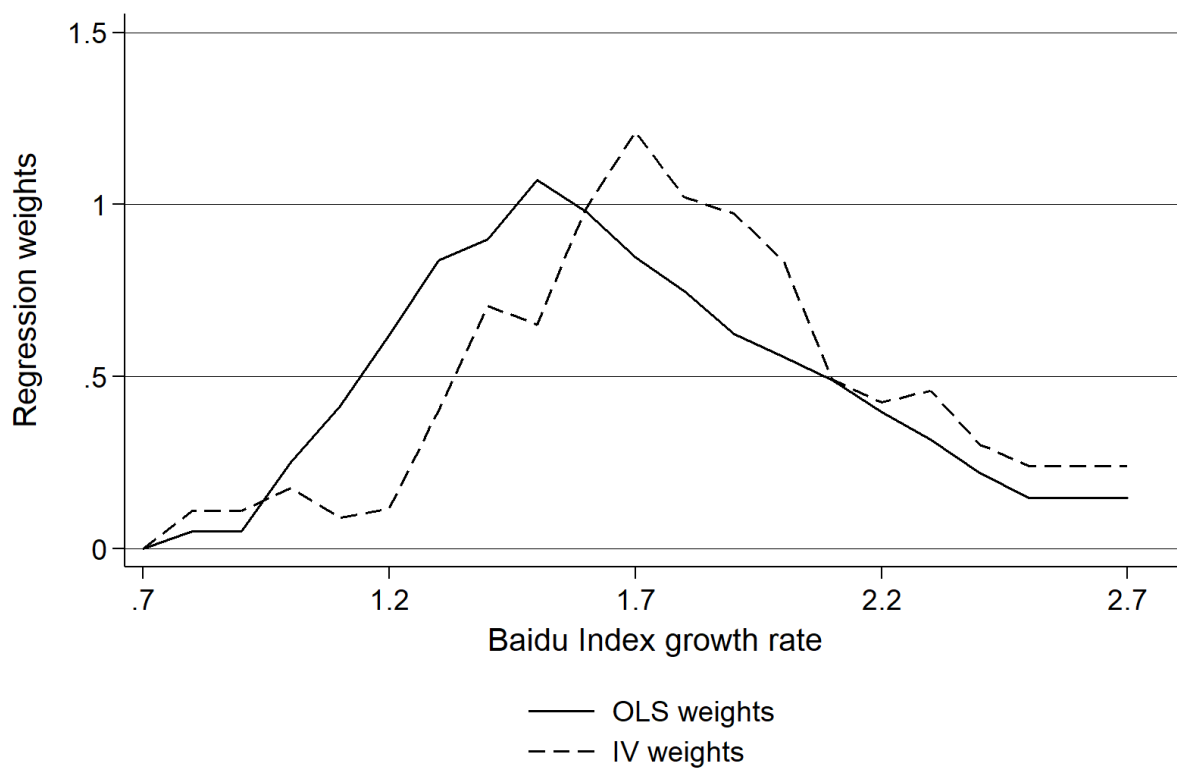
Note: The method to produce the regression weights follows Løken et al. (2012).

Figure 7: The welfare effects of e-commerce

Note: This Figure illustrates the congestion relief effect due to the increase in online shopping. The blue downward-sloping curve is the travel demand function (willingness to pay for various quantities of trips). The demand curve shifts inward to the orange line when e-commerce reduces shopping vehicle demand. The green upward-sloping curve is the average cost of trips. The per-trip cost increases with the number of vehicles on roads due to traffic congestion when there are enough vehicles so that the traveling speed is below free-flow speed. Given that I focus on peak hours, I assume that the average costs of trips increase with the number of trips. The original equilibrium point is (q_1, p_1). The total welfare of consumers from these travels is $A + B$. The welfare loss due to negative externality from traffic congestion is $C + D + E$. The changes in travel demand result in a new equilibrium at (q_2, p_2). The total welfare of consumers is $A + C$, and the welfare loss reduces to E. The welfare gain is then, $C - B$.

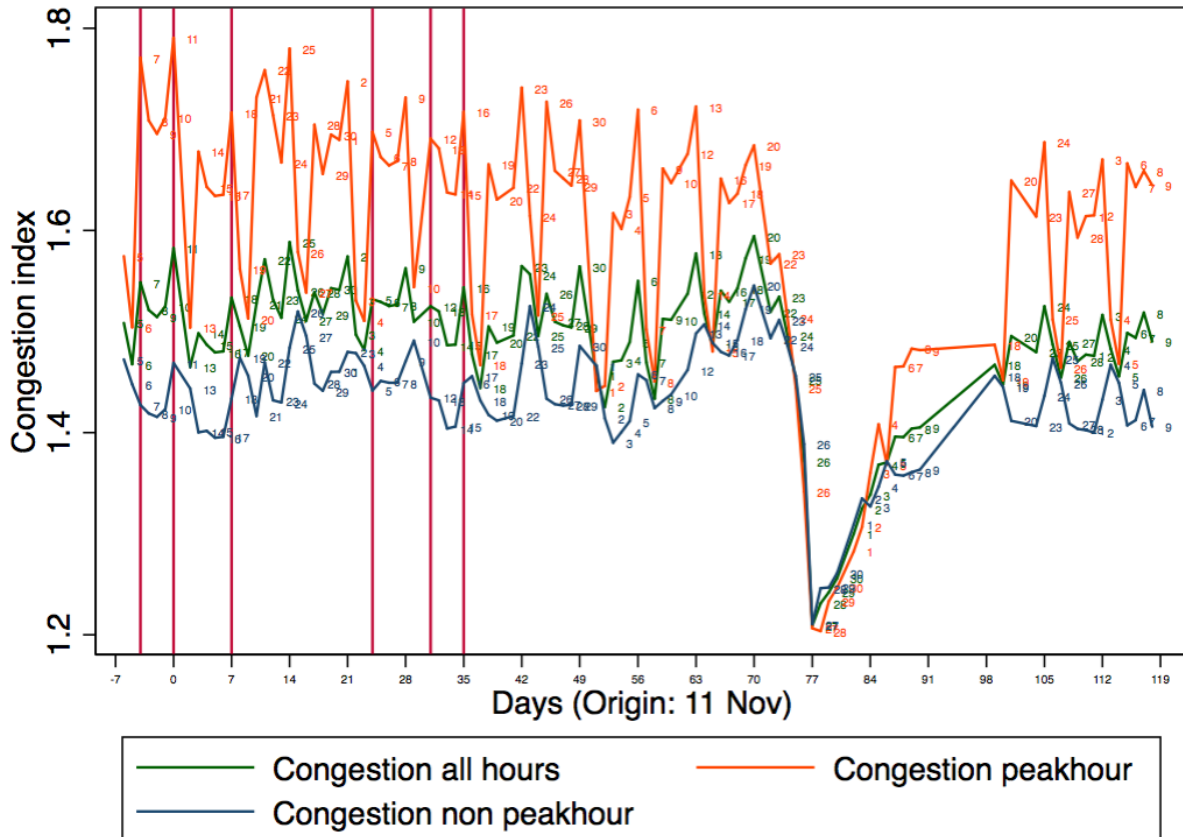# A   Supplementary Figures

Figure 8: The average daily level of traffic congestion

Note: Figure shows the daily average of traffic congestion, peak hour traffic congestion, and off-peak hour traffic congestion from 4 Nov 2016 to 9 March 2017. 0 in the x-axis marks Single Day Shopping Event day. The annotated numbers along the lines are dates of the observation. The first three vertical lines mark the week before and the week after the online shopping event day. The second three vertical lines mark the week before and the week after the offline shopping event day. The plummet of the index in the right shows traffic during Chinese New Year. The traffic congestion index is calculated using average actual travel time to free flow travel time of millions of GPS trails collected map navigation company.
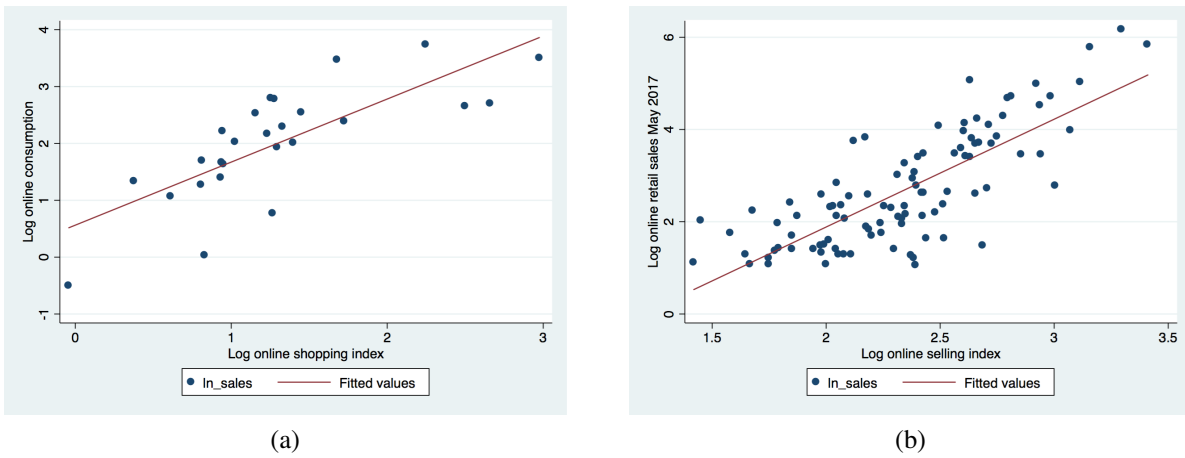
Figure 9: Validation of e-commerce indices

Note: Figure (a) shows that log online consumption during Singles' Day Shopping event day is predicted by the online shopping index. Online consumption data is released by Alibaba and aggregated to province level. The online shopping index is the average of the province. Figure (b) shows that log online selling index is a predictor of log online sales. Online sales data is a overall predicted online sale in May 2017, released by an independent e-commerce research institute.

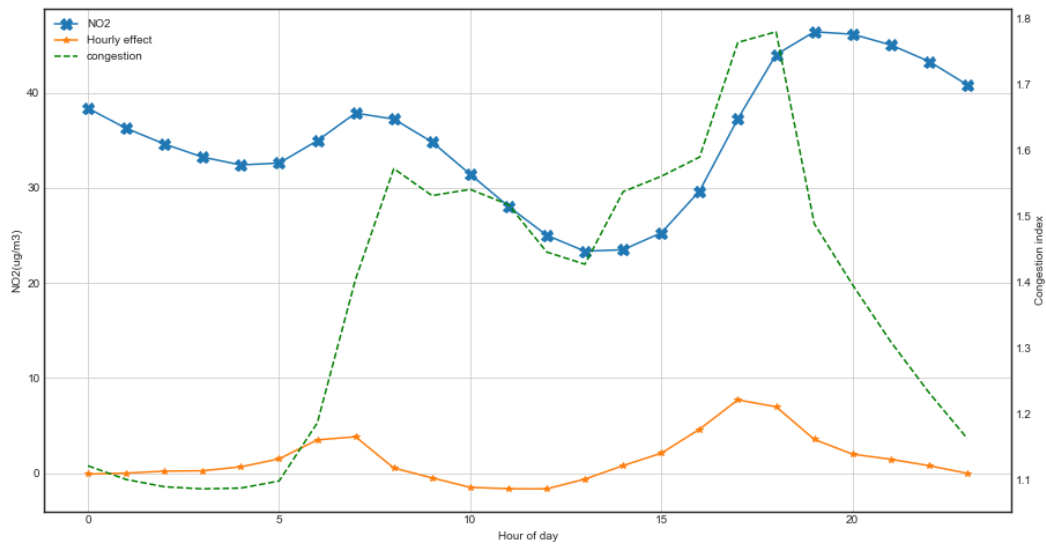Figure 10: Validation of the Baidu index and the online shopping index
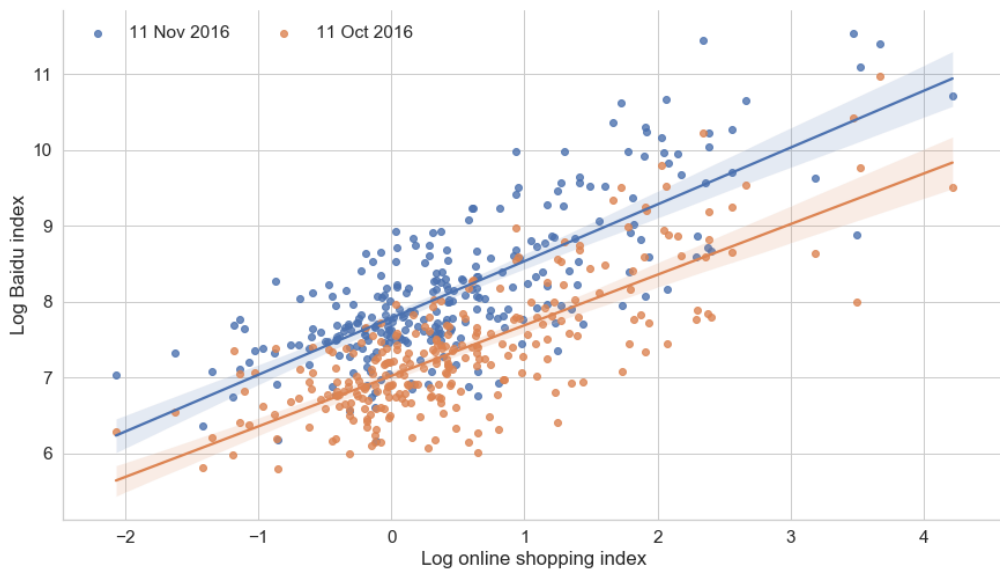
Figure 11: Correlation between level of NO2 and traffic congestion in a day

Note: The figure overlays hourly NO2 shock and cumulative NO2 concentration with the traffic congestion index. Hourly NO2 shock is derived through solving a system of equations as explained in the text. The left Y-axis shows the concentration of NO2 in ug/m3 and the right Y-axis shows the traffic congestion index.
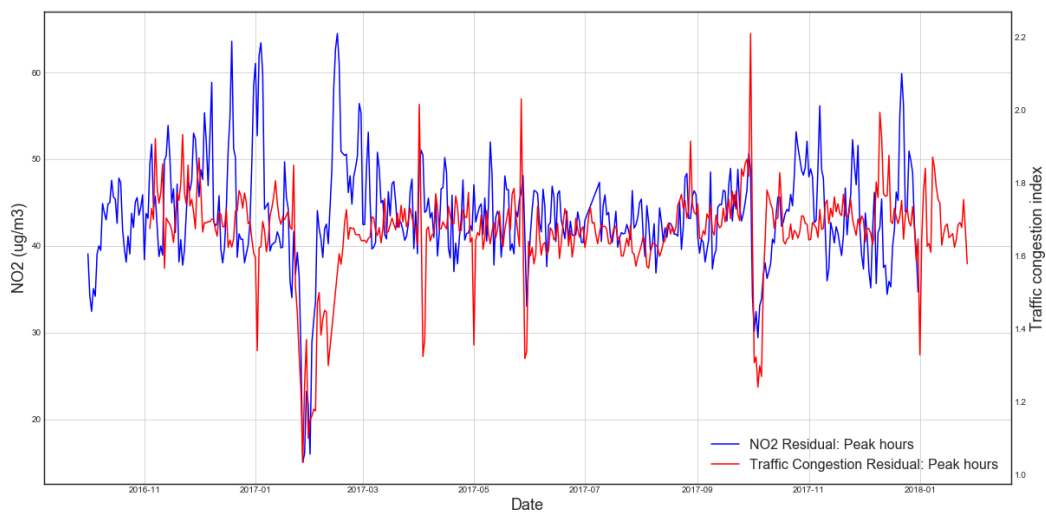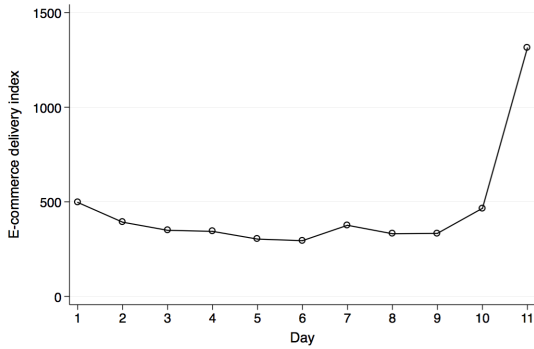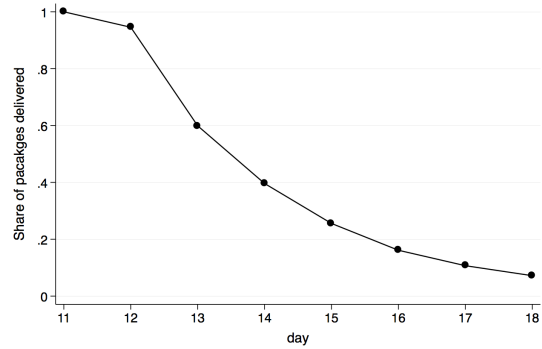

Figure 12: Correlation between level of NO2 and traffic congestion across days

Note: The figure overlays smoothed daily NO2 level of concentration with smoothed traffic congestion index between Nov 2016 and Jan 2018. Both series smoothed by regressing on year, month and day-of-week dummies.

| (a) | (b) |

Figure 13: Time span of online shopping delivery

Note: Figure (a) shows the number of packages delivered before the Singles' Day Shopping Event. Figure (b) shows the time span of online delivery after the Singles' Day Shopping Event
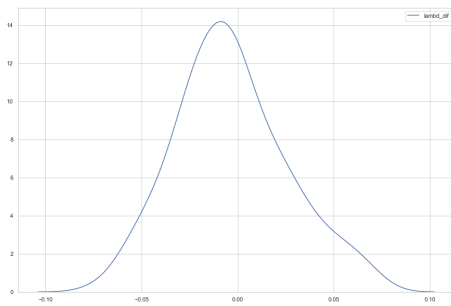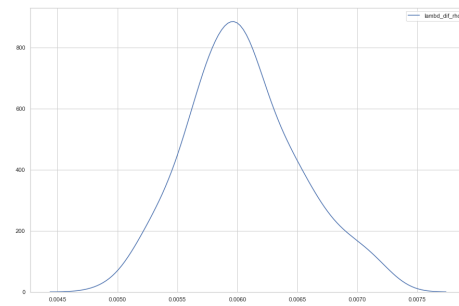
| (a) | (b) |

Figure 14: Simulation results for a validation of the formula for λ

Note: Figure (a) show the distribution of the percentage difference between the λ^{theory} and $\lambda^{simulated}$. Figure (b) shows the distribution of the percentage difference between the λ^{ρ} and $\lambda^{simulated}$.

B   Construction of E-Commerce Indices

$$Online\ Shopping\ Index_i = E-Consumer\ Density\ Index_i \times 0.6$$
$$+Average\ Online\ Consumption\ Index_i \times 0.4$$

$$E-Consumer\ Density\ Index_i = \frac{Density\ of\ E-Consumer_i}{Expected\ Density\ of\ E-Consumer}$$

$$= \frac{Count\ of\ E-Consumer_i \div Population_i}{63.26\%}$$

$$Average\ Online\ Consumption\ Index_i = \frac{Average\ Online\ Consumption_i}{Expected\ Average\ Online\ Consumption}$$

$$= \frac{Average\ Online\ Consumption_i}{32.5\ thousand\ RMB}$$

$$Online\ Selling\ Index_i = E-Seller\ Density\ Index_i \times 0.6$$
$$+Average\ Online\ Sale\ Index_i \times 0.4$$

$$E-Seller\ Density\ Index_i = \frac{Density\ of\ E-Seller_i}{Expected\ Density\ of\ E-Seller}$$

$$= \frac{Count\ of\ E-Consumer_i \div Population_i}{13.47\%}$$

$$Average\ Online\ Sale\ Index_i = \frac{Average\ Online\ Sale_i}{Expected\ Average\ Online\ Sale}$$

$$= \frac{Average\ Online\ Sale_i}{1\ million\ RMB}$$

C   Mathematical Derivations

C.0.1   Derivation of the Elasticity of Online Consumption to the Relative Price Between Two Channels

Given equation 20, and the mean effect is γ times of the share effect, the derivative of online consumption quantity from city i in city j to the price shock can be written as $\gamma + 1$ times the share effect. I take advantage of $k = 1$, which is the key assumption for the result in equation 23 to hold, before the event to simplify the derivatives calculation.

$$\frac{dQ_{oij}}{dk}\Big|_{k=1} = (\underbrace{\pi_o \frac{dA}{dk}}_{\gamma v} + \underbrace{A \frac{d\pi_o}{dk}}_{v})C_{ij} \tag{C1}$$

$$= (\frac{\sigma - 1}{\theta}s^\theta + 1)\frac{d\pi_o}{dk}AC_{ij} \tag{C2}$$

Summing across all origin cities i,

$$\frac{d\sum_i Q_{oij}}{dk}\Big|_{k=1} = (\pi_o \frac{dA}{dk} + A\frac{d\pi_o}{dk})C_{ij} \tag{C3}$$

$$= (\frac{\sigma - 1}{\theta}s^\theta + 1)\frac{d\pi_o}{dk}A\sum_i C_{ij} \tag{C4}$$

which gives the derivative of online consumption quantity in city j from all source cities including itself,

$$\frac{dQ_{oj}}{dk}|_{k=1} = (\pi_o \frac{dA}{dk} + A\frac{d\pi_o}{dk})C_{ij} \tag{C5}$$

$$= (\frac{\sigma - 1}{\theta}s^\theta + 1)\frac{d\pi_o}{dk}A\sum_i C_{ij} \tag{C6}$$

Using equation 11, $A\sum_i C_{ij} = Q_j$ and $Q_{oi} = \pi_o Q_j$, equation C7 can be written as,

$$\frac{dQ_{oj}}{Q_{oj}}|_{k=1} = (\frac{\sigma - 1}{\theta}s^\theta + 1)\frac{d\pi_o}{dk}\frac{1}{\pi_o}dk \tag{C7}$$

Given the derivative of the share of online shopping on price shock is,

$$\frac{d\pi_o}{dk}|_{k=1} = \frac{\theta s^\theta}{(s^\theta + 1)^2}\frac{\sigma}{\sigma - 1} \tag{C8}$$

The growth rate of online shopping can be simplified to,

$$\frac{dQ_{oj}}{Q_{oj}}|_{k=1} = dk\frac{(\sigma - 1)s^\theta + \theta}{s^\theta + 1}\frac{\sigma}{\sigma - 1} \tag{C9}$$

Given that $\frac{\Delta Q_{oj}}{Q_{oj}}/\frac{\Delta k}{k}$ approximates the elasticity of the online consumption quantity to price shock when Δk is small, I obtain,

$$\rho|_{k=1} = \frac{(\sigma - 1)s^\theta + \theta}{s^\theta + 1}\frac{\sigma}{\sigma - 1} \tag{C10}$$

Plugging equation C10 into equation equation 22 gives,

$$\lambda|_{k=1} = -\frac{\theta + (\sigma - 1)s^\theta}{\theta - \sigma + 1} \tag{C11}$$

$$= -\frac{\frac{\sigma-1}{\sigma}\rho(s^\theta + 1)}{\frac{\sigma-1}{\sigma}\rho(s^\theta + 1) - (\sigma - 1)s^\theta - \sigma + 1} \tag{C12}$$

$$= -\frac{\frac{\sigma-1}{\sigma}\rho}{\frac{\sigma-1}{\sigma}\rho - \sigma + 1} \tag{C13}$$

$$= -\frac{\rho}{\rho - \sigma} \tag{C14}$$

Taking the inverse of λ and the absolute value gives,

$$\frac{1}{|\lambda|} = 1 - \frac{\sigma}{\rho} \tag{C15}$$

C.0.2   Simulation Procedure and Results

One of the key predictions from the model is that the online-offline substitution is

$$\lambda^{theory} = \frac{\sum_i \Delta Q_{oij}}{\sum_i \Delta Q_{fij}} = -\frac{\theta + (\sigma - 1)k^\theta s^\theta}{\theta - \sigma + 1}$$

The exercise below verifies this result.

First, I set the number of products as 1000, and the number of consumers as 4000. For simplicity, I assume there are two cities, with one city having 1600 consumers and the other having 2400 consumers. I set that one city produces 300 types of products and the other city produces 700 types of products, which are sold in both cities. Denote vector $\vec{x}$ as the values of city characteristics x in cities i and j. I set hourly wage $\vec{w} = (10, 20)$ and productivity $\vec{p} = (10, 20)$. I set intercity transportation cost $\tau = 1$ if the firm and the consumer are in the same city, and set $\tau = 2$ if the firm and the consumer are in the different cities. Income in each city is assumed to be $\vec{w} \times 40$, assuming workers work 40 hours a week. Consumption quantity is thus calculated on a weekly basis. I create two set of draws of z_o or z_f for each pair of product and consumer, from two Fréchet distributions with an online channel preference parameter $s_o = 0.8$ and an offline channel preference parameter $s_f = 1$. The shape parameter is set as the same in both channels, $\theta = 9$. I set the elasticity of substitution between varieties $\sigma = 4$ following the trade literature. These results hold for other sets of assumed values.

In normal days without a sales event, I set $k_o = k_f = 1$. For a specific variety, I assign consumers who satisfy the condition $k_o z_o > k_f z_f$ to online type and assign the rest to offline type. I then calculate the consumption quantity based on equation 3 for online type and offline type, respectively. The price index is calculated based on equation 4.

In the event, I set $k_o^{event} = 1.1$ while maintaining the values of initial draws of z_m. Based on the old condition $k_o z_o > k_f z_f$ and a new condition $k_o^{event} z_o > k_f z_f$, there are three types of consumers for each variety: consumers that remain online, consumers that switch from offline to online, and consumers that remain offline. Consumers switch from offline to online because for them $k_o z_o < k_f z_f$ while $k_o^{event} z_o > k_f z_f$. Then, I recalculate the consumption quantity for each consumer.

I sum the consumption quantity for online and offline consumers, respectively, to obtain Q_{mj} and Q_{mj}^{event} for city j, which gives the ΔQ_{mj}. I use them to calculate $\lambda^{simulated}$ and compare it with λ^{theory}. I construct a statistic $\frac{\lambda^{simulated} - \lambda^{theory}}{\lambda^{theory}}$ to measure the relative difference between the two

values. Given that $\lambda^{simulated}$ is a random variable, I replicate the above procedures 100 times and plot the distribution of the statistic in Figure 14 (a).

Similarly, I calculate $\rho_j = \frac{\Delta Q_{oj}}{Q_{oj}}$ to obtain the online-offline substitution λ^ρ following equation 23. Again, I construct a statistic $\frac{\lambda^{simulated} - \lambda^\rho}{\lambda^\rho}$ to measure the relative difference between the two values. I replicate the above procedures 100 times and plot the distribution of the statistic in Figure 14 (b).

D Additional Analysis and Results

D.1 The Characteristics of the Cities That Contribute Higher Variations to the IV Estimates

A heuristic approach would be looking at heterogeneous effects of the IV on the endogenous variable in the first stage. Table 11 shows the result when adding interaction of the IV with city characteristics. Column 1 presents the results without any interaction terms as the benchmark. The rest of the columns add the interaction term with the variable specified in the column title. The effect of postage fee on the change of online shopping clearly increases with market potential, income, and the number of online consumers. This suggests that the IV estimates are identified from the cities with such attributes.

Table 11: Heterogeneity in the first-stage

	(1) No interaction	(2) Market potential	(3) Income	(4) Internet	(5) Mobile
Log avg postage fee	0.612***	0.132	0.000	-0.031	-0.014
	(0.223)	(0.134)	(0.008)	(0.038)	(0.028)
Interaction of postage fee with the variable in column title		0.471***	0.088***	0.194***	0.152***
		(0.047)	(0.001)	(0.004)	(0.003)
Market potential	-0.017	-0.319***	0.005	0.024	0.013
	(0.115)	(0.112)	(0.004)	(0.017)	(0.012)
Log GDP per capita	-0.060**	-0.044**	-0.089***	-0.007*	-0.005
	(0.028)	(0.019)	(0.005)	(0.004)	(0.003)
Log internet users	0.016	-0.001	0.000	-0.193***	-0.007
	(0.043)	(0.023)	(0.002)	(0.007)	(0.005)
Log mobile users	0.078*	0.014	-0.001	0.004	-0.137***
	(0.045)	(0.022)	(0.002)	(0.006)	(0.006)
R^2	.38	.79	1	.99	.99
N	90	90	90	90	90

Note: The dependent variable is the growth rate of the Baidu index. The second row of each column shows the coefficients of the interaction term of the log average postage fee with the variable in the column title. Robust standard errors are applied.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

D.2 The Changes in Intercity Traffic

Table 12 quantifies the change in intercity traffic following equation 36. The outcome variable of interest is the travel time in minutes for any routes between cities. The regression sample contains Thursdays and Fridays in the weeks before and after the event. All regressions control for route length. As shown in the first row of the table, the estimated average travel time per km is stable at 0.7 minutes (or 86 kmh) during peak hours and 0.69 during off-peak hours. Intercity roads congestion appeared not to vary much with peak hours. The first three columns report the results for peak hours and the last three columns report the results for off-peak hours. Columns 1 and 4 report the average change in traffic in the first week γ_1 and that in the second week γ_2 for peak hours and off-peak hours, respectively. Column 1 shows that a typical traveler is slowed down by about 24 minutes in peak hours in the first week after the large-scale online sale. The effect continues in the second week, though nearly halved in magnitude. The increase in traffic in the second week is only a third of the first week for off-peak hours. As expected, the delay is more significant during peak hours. Columns 2 and 4 show the results for the interaction specifications, which explores how the characteristics of the route's origin city affect travel time. Cities with a higher online selling index have a bigger spike in traffic congestion, while cities with a higher online shopping index experience, comparatively, less increase in traffic. A possible explanation is that the cities that are involved in more online selling activity may ship more goods out of the city, with the traffic on the intercity roads originating from these cities expected to be more congested. The effect is more salient in the second week. Similar patterns are observed for off-peak hours. Columns 3 and 6 add in the interaction terms of the online shopping index and the online selling index in the destination cities. The characteristics of the route destination cities appear irrelevant. It is desirable to have a longer time series to pin down the diminishing trend of traffic congestion after the event. Unfortunately, I only have the data for three weeks[67]. Nevertheless, this regressive pattern in the post-event travel time is in line with the e-commerce delivery index and share of packages delivered by days surrounding the event, as shown in Figure 13. The pattern of the change in travel time emerging from the three weeks in the 9,126 routes between cities provides strong evidence of the impact of the online sale on traffic congestion on intercity roads.

---

[67]I only employed the Baidu map API for Thursday and Friday in three weeks due to budget constraints because of data collection cost.

Table 12: Estimates of the changes in intercity travel time before and after the event

	(1) Peak hour	(2) Peak hour	(3) Peak hour	(4) Non peak hour	(5) non peak hour	(6) Non peak hour
Distance	0.70***	0.70***	0.70***	0.69***	0.69***	0.69***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Week 1	24.06***	29.35	42.11	27.60***	19.82	24.97
	(8.83)	(44.07)	(51.37)	(7.84)	(38.16)	(41.87)
Week 2	7.84**	-12.06	2.30	11.83***	-2.25	4.87
	(3.16)	(18.29)	(24.74)	(2.24)	(12.22)	(17.77)
Week 1 × Online Selling Origin		3.49	3.43		7.02	7.00
		(22.09)	(22.10)		(19.42)	(19.42)
Week 2 × Online Selling Origin		19.23*	19.15*		10.67	10.63
		(10.44)	(10.45)		(6.96)	(6.97)
Week 1 × Online Shopping Origin		-9.75	-9.76		-6.23	-6.24
		(8.31)	(8.31)		(7.34)	(7.34)
Week 2 × Online Shopping Origin		-18.12***	-18.11***		-7.84**	-7.84**
		(5.42)	(5.42)		(3.45)	(3.46)
Week 1 × Online Selling Destination			-4.66			-1.62
			(12.99)			(8.13)
Week 2 × Online Selling Destination			-6.33			-3.45
			(9.10)			(7.36)
Week 1 × Online Shopping Destination			-1.25			-0.95
			(5.35)			(3.39)
Week 2 × Online Shopping Destination			0.41			0.73
			(4.18)			(3.21)
R^2	0.96	0.96	0.96	0.98	0.98	0.98
N	263302	263302	263302	680925	680925	680925

Note: The dependent variable is travel time in minutes. The reference group is the week before the Singles' Day shopping event. Columns 1-3 show the results for peak hours, and columns 4-6 show results for off-peak hours. Origin city fixed effects, destination city fixed effects, day-of-week fixed effects, and hour fixed effects are included. Standard errors are clustered at the origin cities × destination cities level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

References

Adler, Martin W, Federica Liberini, Antonio Russo, and Jos N van Ommeren, "Road Congestion and Public Transit," ITEA Conference Working Paper 2017.

Akbar, Prottoy A, Victor Couture, Gilles Duranton, and Adam Storeygard, "Mobility and Congestion in Urban India," Working Paper 25218, National Bureau of Economic Research November 2018.

Anderson, Michael L., "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion," *American Economic Review*, September 2014, *104* (9), 2763–96.

Angrist, Joshua D. and Alan B. Krueger, "Chapter 23 - Empirical Strategies in Labor Economics," in Orley C. Ashenfelter and David Card, eds., *Orley C. Ashenfelter and David Card, eds.*, Vol. 3 of *Handbook of Labor Economics*, Elsevier, 1999, pp. 1277 – 1366.

Antràs, Pol, Teresa C. Fort, and Felix Tintelnot, "The Margins of Global Sourcing: Theory and Evidence from U.S. Firms," *American Economic Review*, 2017 2017, *107* (9), 2514–64.

Armington, Paul S., "A Theory of Demand for Products Distinguished by Place of Production (Une théorie de la demande de produits différenciés d'après leur origine) (Una teoría de la demanda de productos distinguiéndolos según el lugar de producción)," *Staff Papers (International Monetary Fund)*, 1969, *16* (1), 159–178.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

Braithwaite, Alan, "The Implications of Internet Shopping Growth on the Van Fleet and Traffic Activity," 2017, (May), 1–6.

Brilon, Werner and Jan Lohoff, "Speed-flow Models for Freeways," *Procedia - Social and Behavioral Sciences*, 2011, *16*, 26 – 36. 6th International Symposium on Highway Capacity and Quality of Service.

Cairns, Sally, "Delivering supermarket shopping: More or less traffic?," *Transport Reviews*, 2005, *25* (1), 51–84.

_ , Lynn Sloman, Carey Newson, Jillian Anable, Alistair Kirkbride, and Phil Goodwin, "Smarter Choices-Changing the Way We Travel Final report of the research project: 'The influence of soft factor interventions on travel demand' Disclaimer and copyright Accompanying report," *Final report of the research project: 'The influence of soft factor interventions on travel demand'*, 2004.

Cavallo, Alberto, "Are Online and Offline Prices Similar? Evidence from Large Multi-channel Retailers," *American Economic Review*, January 2017, *107* (1), 283–303.

Choi, Hyunyoung and Hal Varian, "Predicting the Present with Google Trends," *Economic Record*, 2012, *88* (SUPPL.1), 2–9.

Cooper, Zack, Stephen Gibbons, Simon Jones, and Alistair Mcguire, "Does hospital competition save lives? Evidence from the English NHS patient choice reforms," *Economic Journal*, 2011, *121* (554), 228–260.

Costinot, Arnaud, Dave Donaldson, Margaret Kyle, and Heidi Williams, "The More We Die, The More We Sell? A Simple Test of the Home-Market Effect," Working Paper 22538, National Bureau of Economic Research August 2016.

Couture, Victor, Benjamin Faber, Yizhen Gu, and Lizhi Liu, "E-Commerce Integration and Economic Development: Evidence from China," Working Paper 24384, National Bureau of Economic Research March 2018.

Coşar, A. Kerem, Paul L.E. Grieco, Shengyu Li, and Felix Tintelnot, "What drives home market advantage?," *Journal of International Economics*, 2018, *110*, 135 – 150.

Dolfen, Paul, Liran Einav, Peter J Klenow, Benjamin Klopack, Jonathan D Levin, Laurence Levin, and Wayne Best, "Assessing the Gains from E-Commerce," Working Paper 25610, National Bureau of Economic Research February 2019.

Duranton, Gilles and Matthew A. Turner, "The Fundamental Law of Road Congestion: Evidence from US Cities," *American Economic Review*, October 2011, *101* (6), 2616–52.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Else, P. K., "A Reformulation of the Theory of Optimal Congestion Taxes," *Journal of Transport Economics and Policy*, 1981, *15* (3), 217–232.

Fan, Jingting, Lixin Tang, Weiming Zhu, and Ben Zou, "The Alibaba effect: Spatial consumption inequality and the welfare gains from e-commerce," *Journal of International Economics*, 2018, *114*, 203 – 220.

Forman, Chris, Anindya Ghose, and Avi Goldfarb, "Competition Between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live," *Management Science*, 2009, *55* (1), 47–57.

Gabaix, Xavier, David Laibson, Deyuan Li, Hongyi Li, Sidney Resnick, and Casper G. de Vries, "The impact of competition on prices with numerous firms," *Journal of Economic Theory*, 2016, *165*, 1 – 24.

Gendron-Carrier, Nicolas, Marco Gonzalez-Navarro, Stefano Polloni, and Matthew A Turner, "Subways and Urban Air Pollution," Working Paper 24183, National Bureau of Economic Research January 2018.

Gibbons, Stephen, Teemu Lyytikäinen, Henry G Overman, and Rosa Sanchis-guarner, "New road infrastructure : The effects on firms," *Journal of Urban Economics*, 2019, *110* (March 2017), 35–50.

Goldfarb, Avi and Catherine Tucker, "Digital Economics," *Journal of Economic Literature*, March 2019, *57* (1), 3–43.

Goldmanis, Maris, Ali Hortaçsu, Chad Syverson, and Onsel Emre, "E-commerce and the Market Structure of Retail: E-commerce and Market Structure," *Economic Journal*, 2010, *120* (October), 651–682.

Goolsbee, Austan and Peter Klenow, "Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce," 2018, (January 2014), 488–492.

Gu, Yizhen, Chang Jiang, Junfu Zhang, and Ben Zou, "Subways and Road Congestion," *Working paper*, 2019.

Hall, Jonathan V, John J Horton, and Daniel T Knoepfle, "Pricing Efficiently in Designed Markets: The Case of Ride-Sharing," 2019.

Heckman, James J., Sergio Urzua, and Edward Vytlacil, "Understanding Instrumental Variables in Models with Essential Heterogeneity," *The Review of Economics and Statistics*, August 2006, *88* (3), 389–432.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple, "Export Versus FDI with Heterogeneous Firms," *American Economic Review*, March 2004, *94* (1), 300–316.

Henderson, J Vernon, "Effects of Air Quality Regulation," *The American Economic Review*, 1996, *86* (4), 789–813.

Kortum, Samuel S., "Research, Patenting, and Technological Change," *Econometrica*, 1997, *65* (6), 1389–1419.

Krugman, P., "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 1980, *70* (5), 950–959.

Løken, Katrine V., Magne Mogstad, and Matthew Wiswall, "What Linear Estimators Miss: The Effects of Family Income on Child Outcomes," *American Economic Journal: Applied Economics*, April 2012, *4* (2), 1–35.

McFadden, D., "Conditional Logit Analysis of Qualitative Choice Behaviour," in P. Zarembka, ed., *Frontiers in Econometrics*, New York, NY, USA: Academic Press New York, 1973, pp. 105–142.

Notley, S.O., N. Bourne, and N. B. Taylor, "Speed, flow and density of motorway traffic," TRL Insight Report INS003, Crowthorne: Transport Research Laboratory. 2009.

Parry, By Ian W H and Kenneth A Small, "Should Urban Transit Subsidies Be Reduced?," *American Economic Review*, 2009, *20036*, 700–724.

Punakivi, Mikko, "Comparing alternative home delivery models for e-grocery business." PhD dissertation, Helsinki University of Technology 2003.

Redding, Stephen J., "Goods Trade, Factor Mobility and Welfare," *Journal of International Economics*, 2016, *101*, 148 – 167.

_ **and Daniel M. Sturm**, "The Costs of Remoteness: Evidence from German Division and Reunification," *American Economic Review*, December 2008, *98* (5), 1766–97.

_ **and David E Weinstein**, "Measuring Aggregate Price Indexes with Demand Shocks: Theory and Evidence for CES Preferences," Working Paper 22479, National Bureau of Economic Research August 2016.

Relihan, Lindsay E, "Is online retail killing coffee shops?," Working paper 2017.

Seim, Katja and Michael Sinkinson, "Mixed pricing in online marketplaces," *Quantitative Marketing and Economics*, Jun 2016, *14* (2), 129–155.

Tikoudis, Ioannis, Erik T Verhoef, and Jos N Van Ommeren, "On revenue recycling and the welfare effects of second-best congestion pricing in a monocentric city," *Journal of Urban Economics*, 2015, *89*, 32–47.

Underwood, R. T., *Speed, Volume and Density Relationships*, Yale University Bureau of Highway Traffic, 1961.

Varian, Hal R., "A Model of Sales," *The American Economic Review*, 1980, *70* (4), 651–659.

Vosen, Simeon and Torsten Schmidt, "Forecasting Private Consumption Survey Based Indicators VS Google Trends," *Journal of Forecasting*, 2011, *30* (6), 565–578.

Walters, A. A., "The Theory and Measurement of Private and Social Cost of Highway Congestion," *Econometrica*, 1961, *29* (4), 676–699.

Yang, Jun, Avralt od Purevjav, and Shanjun Li, "The Marginal Cost of Traffic Congestion and Road Pricing : Evidence from a Natural Experiment in Beijing," *American Economic Journal Economic Policy (Forthcoming)*, 2018.

__ , **Shuai Chen, Ping Qin, Fangwen Lu, and Antung A Liu**, "The Effect of Subway Expansions on Vehicle Congestion : Evidence from Beijing," *Journal of Environmental Economics and Management*, 2018, *88*, 114–133.

__ , **Ying Liu, Ping Qin, and Antung A Liu**, "A Review of Beijing's Vehicle Registration Lottery Short Term Effects on Vehicle Growth and Fuel Consumption," *Energy Policy*, 2014, *75*, 157– 166.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1645	Dan Andrews Chiara Criscuolo Peter N. Gal	The Best versus the Rest: Divergence across Firms during the Global Productivity Slowdown
1644	Christopher Cornwell Ian M. Schmutte Daniela Scur	Building a Productive Workforce: The Role of Structured Management Practices
1643	Paul Dolan Georgios Kavetsos Christian Krekel Dimitris Mavridis Robert Metcalfe Claudia Senik Stefan Szymanski Nicolas R. Ziebarth	Quantifying the Intangible Impact of the Olympics Using Subjective Well-Being Data
1642	Xavier Jaravel Erick Sager	What are the Price Effects of Trade? Evidence from the US and Implications for Quantitative Trade Models
1641	Johannes Boehm Jan Sonntag	Vertical Integration and Foreclosure: Evidence from Production Network Data
1640	Teodora Borota Fabrice Defever Giammario Impullitti	Innovation Union: Costs and Benefits of Innovation Policy Coordination
1639	Monica Langella Alan Manning	Residential Mobility and Unemployment in the UK
1638	Christos Genakos Mario Pagliero	Competition and Pass-Through: Evidence from Isolated Markets
1637	Holger Breinlich Elsa Leromain Dennis Novy Thomas Sampson	Voting With Their Money: Brexit and Outward Investment by UK Firms

1636	Maria Sanchez-Vidal	Retail Shocks and City Structure
1635	Felipe Carozzi Sefi Roth	Dirty Density: Air Quality and the Density of American Cities
1634	Nicholas Bloom John Van Reenen Heidi Williams	A Toolkit of Policies to Promote Innovation
1633	Stephan E. Maurer Ferdinand Rauch	Economic Geography Aspects of the Panama Canal
1632	Nikhil Datta	Willing to Pay for Security: A Discrete Choice Experiment to Analyse Labour Supply Preferences
1631	Gabriel M. Ahlfeldt Volker Nitsch Nicolai Wendland	Ease Versus Noise: Long-Run Changes in the Value of Transport (Dis)amenities
1630	Grace Lordan Alistair McGuire	Widening the High School Curriculum to Include Soft Skill Training: Impacts on Health, Behaviour, Emotional Wellbeing and Occupational Aspirations
1629	Per-Anders Edin Tiernan Evans Georg Graetz Sofia Hernnäs Guy Michaels	Individual Consequences of Occupational Decline
1628	Pawel Bukowski Filip Novokmet	Between Communism and Capitalism: Long-Term Inequality in Poland, 1892-2015
1627	Thomas Sampson	Technology Gaps, Trade and Income