

Rational pricing in prostitution: Evidence from online sex ads *

Gregory DeAngelo[†] Jacob N. Shapiro[‡] Jeffrey Borowitz[§] Michael Cafarella[¶]
Christopher Ré^{||} Gary Shiffman^{**}

This version August 23, 2018.

Keywords: Illicit Market, Compensating Differentials, Prostitution, Machine Reading

JEL codes: D81, C55, C63, K14

*We thank our anonymous reviewers and participants in seminars at Harvard University and Princeton University for helpful comments and feedback. All errors are our own. This research was supported in part by DARPA. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of DARPA or Giant Oak.

[†]Associate Professor, Claremont Graduate University; E-mail: gregory.deangelo@cgu.edu.

[‡]Professor, Princeton University; E-mail: jns@giantoak.com. Corresponding author.

[§]Social Scientist, Giant Oak,; Email: jeffrey.borowitz@giantoak.com.

[¶]Assistant Professor, University of Michigan; E-mail: michjc@umich.edu.

^{||}Associate Professor, Stanford University; E-mail: chrismre@cs.stanford.edu.

^{**}CEO, Giant Oak; E-mail: gary.shiffman@giantoak.com.

Abstract

The movement of many human interactions to the internet has led to massive volumes of text that contain high-value information for social scientists. Unlocking these texts' scientific value is challenging because they are written for human consumption with a great deal of obfuscation and slang, particularly so in areas of illicit behavior. We use the DEEPDIVE system which conducts large scale probabilist inference in a user-enabled feedback loop to extract structured data from more than 30 million online ads for real-world sex over a four-year period, data significantly larger than that previously developed on the topic. We establish prices in a common numeraire and study the correlates of pricing, focusing on risk. We show first that there is a 15-19% price premium for services performed at a location of the buyer's choosing (outcall). Examining how this premium varies across cities and service venues (i.e. incall vs. outcall) we show that most of the variation in prices is likely driven by supply-side decision making. We decompose the price premium into travel costs (75%) and a remainder that is strongly correlated with local violent crime risk, a proxy for physical risk to the provider. Finally, we show that sex workers demand compensating differentials for the unpleasantness of the job on par with the very riskiest legal jobs; an hour spent with clients is valued at roughly \$151 for incall services compared to an implied travel cost of \$36/hour. These results show that offered prices in the online market for real-world sex are driven by the kinds of rational decision-making common to most pricing decisions and demonstrate the value of applying machine reading technologies to complex online text corpora.

Introduction

The movement of many human interactions to the internet has led to massive volumes of text that contain high-value information for social scientists. For example, online illicit sex markets have yielded tens of millions of sex provider advertisements and over one million customer reviews of those providers. These online texts describe prices, locations, personal characteristics, preferences about the commercial encounter, and other information that is useful for social science but otherwise difficult to obtain at a large scale.¹ Unfortunately, these texts are intended for individual human, not analytical, consumption: they are casually-written and contain nonstandard usage and slang. To date, preparing such data for statistical analysis has required substantial human annotation, with the concomitant expense and necessary reduction in data size.

Recent advances in computer science enable macroscopic analysis of such data at finer resolution than previously possible by extracting high-quality structured analysis-ready information from text and images with minimal human annotation. In this paper we employ the DEEPDIVE system to create a structured database of facts recovered from human-written source texts in the online illicit sex market. DEEPDIVE uses large-scale probabilistic inference in a user-enabled feedback loop, thereby avoiding problems common to most standard annotation techniques, such as reliance on brittle rules or fixed grammars. In a number of applications, DEEPDIVE obtains accuracy that is similar to that of a human annotator Callaway (2015); Peters et al. (2014). We are thus able to obtain a very large and high-quality database about subtle concepts, derived from an extremely messy collection of documents (table A1 provides precision/recall figures).

Data on this market are relatively easy to access in small quantities because much of the activity happens in public web fora (e.g. www.backpage.com). The postings on these sites are text advertisements written in informal English, akin to classified ads, often accompanied by images. As with classified ads, there is informal broad agreement about the kind of information to provide (prices, locations, etc) but diversity in both mode of expression (slang, colloquial usage, nonstandard usage) and in exactly which data values are provided. Figure 1 displays a full example of an online advertisement as well as two examples of ad text to illustrate the linguistic challenges in this space.

¹Other arenas of human interaction producing massive corpora of online text include weapons sales, labor markets, and small-cap stock fraud, among others.

Our text collection contains information from almost 30 million text ads for sex services, scraped from 19 distinct websites between early-2011 and January 2016 by IST Research.

[INSERT FIGURE 1 ABOUT HERE.]

Little is known about the market for sex services, but there are many reasons to want to know more. The sex services market is a two-sided markets where buyers aim to connect with providers of sex services and providers wish to offer their services. Given the illegal nature of the services, both the service provider and John are concerned with the possibility that they could be matching with a law enforcement agent. Additionally, service providers also face a risk of encountering a violent John.² In fact, small scale surveys have found that as many as two in three sex service providers have been assaulted by customers or pimps (Weitzer, 1999).

Sex services were traditionally solicited in outdoor spaces, which resulted in the creation of red-light districts (Hubbard and Sanders, 2003). In the presence of mounting social pressure and threat of arrest, sex service workers were largely relegated to specific locations in urban areas where illegal activities were more tolerated. However, the introduction of the internet fundamentally changed the nature of the initial interaction between clients (the demand side of the market) and service providers (the supply side). Instead of face-to-face interactions, the initial interactions between potential clients and service providers began when a client responded to an advertisement, which offered sex services. The movement to arranging services online versus through face-to-face interactions is thought to result in more safety for service providers (Bass, 2015), as service providers can screen potentially violent clients.³ Additionally, the movement to online advertisements enabled service providers to coordinate appointments and have more control over the location where services are to be performed, rather than waiting at specific outdoor locations or propositioning potential clients in public locations (e.g. bars, casinos). So, service providers can now determine whether they are willing to travel to a potential client's location (outcall), whether the client must come to the location of the service provider's choosing (incall) or if the service provider can accommodate

²As noted in Potterat et al. (2004), the estimated female homicide rate is 204 per 100,000, which is 51 times higher than the second most dangerous occupation for females (liquor store employee).

³A digital footprint of the arrangements of the services to be provided, which can include date, time, location, services to be performed and cost of services, are also generated when arrangements are made by email, which is also thought to deter violent clients.

either situation. Despite these differences in the arrangement of services, payment for services has largely remained unchanged. Cash is still the currency of such transactions, which is most often paid upon completion of services. Violence against the service provider is thought to be most likely to occur at the point of payment, creating a market for additional security.⁴

While product differentiation existed in this market prior to online platforms, it was more difficult for such differentiation to occur. Service providers could place advertisements in alternative weeklies (e.g. *The Village Voice*) or they could work with an escort agency, which would coordinate service providers and potential clients for a fee. The movement to online advertisements enabled sex service providers to become more entrepreneurial and independent, enabling them to keep a greater share of the proceeds from the services that they offered. The movement online also enabled further horizontal and vertical differentiation of services. Vertical product differentiation was enabled, as idiosyncratic preferences for services (e.g. massage, erotic massage, escort, BDSM, “girlfriend experience”, etc.) could be catered to and advertised across providers who are willing to perform such acts. Horizontal product differentiation was also further enabled, as the costs of advertising and searching for one’s ideal variety were both reduced through online platforms. Moreover, advertisements could link to review web sites that enabled a client to see testimonials of the service provider’s quality. Thus, a service provider could generate a reputation, which enabled the service provider to command higher rates for services performed (Cunningham and Kendall, 2016).

The movement of sex service advertisement online also resulted in significantly new knowledge about the market for sex services, as measuring the supply for these services was nearly impossible prior to its movement online. Although crude, we can now measure the number of sex service advertisements by city on a given day. Table 10.4 of Cunningham and Kendall (2011c), for example, reports the average number of sex services advertisements on one platform per MSA population across 31 of the largest municipalities in the US that are offered in a day range from 0.36 (Cleveland) to 18.34 (San Francisco). Unfortunately, such robust measures of the demand for sex services do not exist. In one of the only studies that attempts to estimate the demand for sex services, Roe-Sepowitz et al. (2016) estimate the demand for sex services in 15 large municipalities in the US.

⁴(Dank et al., 2014) provide a very thorough review of market structure and size estimates for eight major American cities.

On average, the study finds that 1 out of every 20 males over the age of 18 years old in these jurisdictions was soliciting online sex ads.

From an academic perspective, the online market for sex is representative of the broader class of markets in which regulation and contract enforcement are decentralized because the underlying activity is illegal.⁵ From a policy perspective, an increasingly large share of commercial sex transactions are coordinated through online markets (Cunningham and Kendall, 2011*b*). The emergence of robust online markets for sex have been associated with a range of social ills including child prostitution (Hughes, 2002; Mitchell et al., 2010), human trafficking (Latonero et al., 2011), and a drop in the average age of prostitutes (Cunningham and Kendall, 2011*b*). At the same time, these markets may reduce transaction costs in the market for sex and enable better use of reputational mechanisms, both of which can be welfare enhancing for buyers and sellers (Cunningham and Kendall, 2011*b*). Cunningham, DeAngelo and Tripp (2018) also notes that the introduction of online sex service clearinghouses (namely, the erotic services section of www.craigslist.com) significantly reduced female homicides. Using the data from these markets to better understand the commercial sex trade therefore has great potential.

Our analysis makes a concrete methodological contribution as well. Because only some online fora are well-structured, and text ads have nonstandard content that is difficult for traditional natural language processing (NLP) methods, past sex market researchers have used relatively small amounts of data.⁶

As a result of these relatively small data sizes, relevant statistics must be aggregated into coarse geographic or temporal regions in order to be statistically useful. For example, a traditional small nationwide sample might yield only a few advertisements from a given city, forcing the analyst to

⁵Such markets include gray markets devoted to goods that are not illegal, but rather are sold second hand not through the original manufacturer, as well as black markets in illegal goods.

⁶Cunningham and Kendall (2011*a*) use data only from www.craigslist.com for select cities, while Cunningham and Shah (2014) use information from www.theeroticreview.com as well as local advertisements in newspapers in Rhode Island. Edlund, Engelberg and Parsons (2009) analyze 40,000 posts from one high-end escort site. Delap (2014) uses data on 190,000 profiles of escorts from a review site where escorts pay to post profiles and are asked to fill in specific fields when setting up their site, making the scraping and analysis relatively straightforward. Moffatt and Peters (2004) use data on 998 complete reports filed by users on one website between January 1999 and July 2000.

aggregate advertisements at a state level in order to retain a minimal number of counts in each aggregated group. In contrast, our extracted database is significantly larger than even the largest previous effort. We extract price/location tuples for 4.5M ads, of which 2.1M occurred in locations for which we have the full set of covariates.⁷ Elements in this large and high-accuracy dataset do not have to be aggregated into very coarse groupings in order to retain statistical validity: the data is *higher-resolution* than past efforts. For example, there may no longer be any need to aggregate advertisements to the state level; many cities will retain sufficient counts. This high resolution data enables us to control for local-level variation in contextual factors (such as local wages or commute times) that would have been impossible with data aggregated at coarser levels.

Using these unique data we find that pricing in the market is broadly rational from an economic perspective. This is not surprising; previous survey-based research has shown that prostitutes charge a premium for risky behaviors and that the size of the premium is greater for more attractive sex workers (Gertler, Shah and Bertozzi, 2005). Exploiting within-period/within-city variation in the pricing structure across different service venues shows that services performed at a location of the buyer’s choosing (so-called ‘outcall’) earn an estimated 18% price premium, approximately \$23 more for an hour-long session, controlling for a wide range of factors.

There could be several reasons for this premium. On the supply side, allowing the buyer to choose the location entails both additional physical risk and additional travel time. On the demand side, customers may be willing to pay a premium to reduce their risk and travel time. To assess the magnitude of these sources of variation in pricing we compare how prices vary as the physical size of the MSA for which services are offered expands and as the rate of violent crime in an area changes. Critically, the difference in those correlations across incall services, i.e. service at the provider’s chosen location, and outcall services, i.e. services at the customer’s chosen location, provides a way to sort out supply from demand elasticities. We find that prices for incall services are uncorrelated with MSA size and violent crime rates once some basic controls are added. Prices for outcall services, however, are strongly positively correlated with MSA size, though they are not correlated with violent crime rates once MSA fixed effects are accounted for.

These results are consistent with the incall and outcall markets being relatively serrated markets.

⁷The full corpus is just under 30M ads, but many lack content on price or define a geographic area that cannot be reliably matched to a region for which other measures are available (i.e. crime rates or employment statistics).

If incall/outcall were one market then we should see prices moving in opposite directions regardless of whether supply is elastic, demand is elastic, or both. That is women living in larger areas who do not want to travel should compensate men to come to them by offering lower prices and they should charge more for travel. That we primarily observe movement in the outcall market across city size suggests that both supply and demand are fairly inelastic with respect to distance in the incall market but not in the outcall market.

The magnitude of the increase for outcall as city size increase indicates a price for providers' travel time of \$36 per extra hour of average commute time in a city, much smaller than the \$151/hour mean price for incall time with a client, but much larger than the average female wage of \$14/hour in our sample.⁸ That difference is consistent with workers in this market demanding substantial compensation to make up for the distastefulness of time spent with clients, a compensating differential very large compared to the differentials that are easily measured.⁹

These results have a number of policy implications. Most importantly, improved labor market opportunities for women appear to change the composition of suppliers in the market but does not necessarily reduce the volume of activity, at least as proxied by ad postings. Secondly, the large difference between compensation required for travel and that for time spent with clients implies that many workers in the market would happily shift to other activities given the opportunity. Finally, with regards to some risks associated with prostitution (i.e. violence), it appears that sex workers advertising online may have sufficient market power to demand compensation for those risks in the outcall market, implying that the supply of workers in that market is inelastic.

The remainder of this paper proceeds as follows. Section 1 provides background on the online market for sex services. Section 2 outlines the technological innovations that enable this research. Section 3 briefly introduces the data. Section 4 analyzes the relationship between pricing and social conditions, including economic opportunities for potential providers. Section 5 concludes.

1 Background

This section provides basic background information on the online market for real-world sex.

⁸95% confidence intervals on the estimated premium ranges from \$11 to \$73.

⁹Lavetti (2015) estimates slightly smaller compensating differentials in fisheries that have exceptionally high annual mortality risks.

1.1 Online ads for sex

Online advertisements for sex contain a wide variety of information, presumably whatever the seller deems necessary to drive demand for their services. The style of ads varies by website, some contain explicit language (e.g., “fetish friendly”, specify sex acts, or clearly discuss prices per hour), while others use more veiled language that almost implies dating (e.g., “Gf services always offered :)”). Ad postings include varying levels of information including age, ethnicity, height, weight, build, hair and eye color, and measurements. Some advertisements include links to the provider’s reviews by their past clients. Almost all advertisements include images, most of which are sexually suggestive or explicit, and phone numbers that allow clients to contact them. Sometimes, ads include guidelines for conduct on the phone with them; examples include ‘no texting’ and ‘no foul language.’ The market is clearly segregated by provider gender and we focus on ads for services by women.

1.2 Ad sources

Online ad content in this market comes from three distinct sources. First, there are individual providers who post ads representing themselves and pay on a per-ad basis. Second, there are content aggregators which repost content from backpage.com or other sites (see. e.g. Escortphonenumber.com) and make money by selling ad space on their websites. Third, there are spammers who post ads in many locations seeking to drive traffic to other websites on which they sell space to advertisers or goods to those who click through (e.g. ‘click here to see my pics’ or redirects to other websites requiring payment for services).¹⁰ In our analysis we focus on the first type by filtering out duplicate ads and by restricting the sample to ads whose contact information is not reused too frequently.

1.3 Differences between sites and users

Content on the 19 sites scraped for this analysis is as varied as in any other online market along two dimension. First, there is site-level variation. Some sites are very formal and employ a standardized format (e.g. theeroticreview.com), while others allow ad hoc posts similar to craigslist.com. The

¹⁰Content aggregators account for roughly 38% of the ads in our corpus. Sometimes these sites include old ads not still present on other sites and so we include them in our collection.

largest website in our sample, backpage.com, had almost 16.8M ads posted between 2013 and early-2016, while the smallest, myredbook.com, had only 35,400 during the same period. The share of ads with prices varied between 2.1% on several sites up to 80.0% of the 35,400 ads scraped from myredbook.com. Most websites covered a wide geographic area, more than 200 unique MSAs per site, but the share of ads with location information varied widely. On average 68% of the ads had extractable location information, but the rate by website varied from a high of 87% for cityvibe.com to a low of only 20% for myproviderguide.com.

Second, there is variation in ad content due to the decisions of providers about how much to disclose. Prior research shows that providers who offer more information command higher prices (Logan and Shah, 2013). Including more information also entails greater risks from law enforcement. The more potentially identifying information a provider offers the easier it is for law enforcement to track them down. While this may be a minor concern for independent voluntary providers, those who are underage or operating in jurisdictions where police pursue prostitution face a clear tradeoff.

2 Technology

High quality extraction from free-form text is challenging because of the massive amount of linguistic variation possible. Standard text processing approaches such as regular expressions are quite brittle and dependent on small changes in the source text. NLP methods are often effective at discovering linguistic information about the text (e.g., parse trees), but do not alone solve the extraction challenge.

DEEPDIVE (Zhang (2015)) is a system for extracting relational databases from unstructured text. It is distinctive when compared to previous information extraction systems in its ability to obtain very high quality databases reasonable engineering cost.¹¹ The system ingests raw documents and emits a structured database.

The most important component of DEEPDIVE is a novel developer framework that allows an engineer to reliably and rapidly improve extracted data quality, until the output database is as good

¹¹The output database has high precision (i.e., it contains facts that very accurately reflect the source documents) and high recall (i.e., most of the facts in the source documents appear in the output database).

as the downstream application requires. Internally, DEEPDIVE includes a high-performance engine for statistical inference and learning, allowing it to handle data that is noisy and imprecise. It is enabled by a number of recent innovations in scalable machine learning and data management (Shin et al. (2015); Zhang and Re (2014); Recht et al. (2011)). Figure 2 outlines the overall DEEPDIVE processing pipeline.

[INSERT FIGURE 2 ABOUT HERE.]

As described in detail in (Zhang (2015)), the system entails a three-phase process. For each phase, the engineer writes a short piece of program code, usually in Python.

- The *candidate generator* is an engineer-written function that is applied to each input document and yields candidate extractions. The goal of the candidate generator is to eliminate “obviously” wrong outputs (e.g., non-numeric prices). Its output should be high-recall, low-precision.
- The *extraction features* are user-defined functions that are applied to each candidate emitted in the previous step. An extraction feature is intended to encode a user-understandable piece of evidence about each candidate, useful when deciding whether the candidate is a correct extraction or not. For example, does a candidate for price have a \$ symbol to its left? Obtaining high accuracy often means using many high-quality extraction features. Unlike some statistical frameworks in which features are largely or entirely synthetic, all DEEPDIVE features are designed to be comprehensible by humans to permit manual debugging.
- The *distant supervision rules* provide a positive or negative label to some of the feature-enriched candidates. For example, there might be extremely unambiguous prices that can be safely labeled as correct extractions. Alternatively, 9-digit zip codes that begin with a non-zero are unambiguously not prices for most applications, and can be safely labeled as incorrect price extractions. Despite the inevitable labeling errors that such rules introduce, we have found this approach to be preferable to the time-consuming process of manually providing labels.

After applying the above three steps, DEEPDIVE constructs a large factor graph model that creates a random boolean variable for each candidate. The system infers a probability for each

candidate, then applies a threshold to each inferred probability to determine whether the extraction will be placed in the output database (e.g., a given candidate is determined to be a one-hour price if the inferred probability exceeds 0.75). In this paper, DEEPDIVE ingests a corpus of unstructured sex ads, then produces a high-quality output database which is used for all of our social science analysis.

3 Data

We analyze two unique data sets in our analysis. The first data set is derived from content in nearly 30 million online ads for sex across 19 different websites. The scrape was conducted as part of the Defense Advanced Research Project Agency’s (DARPA) *Memex* project by IST Research between early 2011 and January 2016. 51.2% of the data come from www.backpage.com, another 10.0% were scraped off sites which repost ads from www.backpage.com and other sources (including www.escortphonenumber.com, www.escortsincollege.com, www.escortads.com, www.massagetroll.com, and www.escortsintheus.com), and the remainder come from various sections of www.craigslist.com which still hosted escort ads at the time of data collection (4.1%) as well as smaller focused sites such as www.cityxguide.com (2.9%), www.myproviderguide.com (2.3%), and a number of smaller sites.

The second data set contains information from approximately 1.1 million online reviews of sex services recorded at the web site www.theeroticreview.com, which was, until recently, the largest website hosting reviews of sex services. These data were scraped in March 2016. Since the website archives old content the reviews are from as early as 1998 through March 2016.

We describe each of these data sets in detail below.

3.1 Advertisement Data

Each ad consists of a free-form text field, and — depending on the site — additional structured fields such as post date or location. When available, these fields are scraped using the HTML structure of the site and merged with other extractions. Most content, however, is only available in the free-form text. Scraped ads which had a specific service location were mapped to a Census Bureau Metropolitan Statistical Area (MSA), the smallest geographic unit for which reliable labor

force data are available in time-series across the United States.

While the initial scrape contains 29.9 million online ads, our final data set is the subset of the initial data for which we could extract information on the full range of relevant covariates. Whether an ad is for incall, outcall or both services is extracted from all ads. Posting date information could not be accurately extracted from approximately 5 million ads. Location was unclear in another 13 million ads (i.e. the scraped text did not specify the locations for which services were offered). Prices charged by providers were missing or unclearly stated in another 8 million ads. Finally, 1.8 million ads with location information were for small towns that could not fit clearly within one of the Census MSA locations. Our final estimation data contain all 2.1M ads for which we have information on the full set of covariates identified below, do not represent ads for providers working in massage parlors, and are not posted at unrealistically high rates.

Table 1 provides summary statistics of the ad-level data for all ads that did not appear to be spam (Panel A), for the 2.1M ads that could both be linked to an MSA for which we have the full set of covariates (Panel B), as well as the MSA/month (Panel C) and MSA level covariates Panel (D).

3.2 Review Data

Review data from www.theeroticreview.com (TER) are broken into two components. First, each service provider has a top page where specific structured text about the provider can be obtained. For example, the provider's age range, hair color, email, phone number, preferences for performing services at a location of their choice or the John's choice, average performance rating, and appearance statistics can be found on this page. Second, on separate subpages the specific reviews from each John that reviewed the provider can be accessed. Each review contains a series of additional information about the provider, including appearance and performance ratings that are specific to that review, whether the service was an incall or outcall, type of intercourse, and precautions taken by the provider such as the use of a screening agency.

Table 2 provides summary statistics for the 1.1 million reviews from TER. Figure 3 shows the average incall price per hour for each of the 82 most populous MSAs in the United States. Circles are sized by ads per capita. Darker shading indicates higher median ad prices within the MSA.

[INSERT FIGURE 3 ABOUT HERE.]

[INSERT TABLES 1 and 2 ABOUT HERE.]

3.3 Control Variables

To analyze the relationship between prices in the ads and other factors we include a broad range of MSA-level covariates from the following sources:

- Opportunity costs. We obtain measures of unemployment from the American Community Survey (ACS) data files for each MSA at an annual level. To assess travel time providers of outcall services can expect, we generate average commute (which is asked as a categorical variable) by weighting the midpoint of response times for each bucket of respondents by the proportion of respondents for that bucket.
- Wages. To control for economic opportunity we generate a Bartik-style wage instrument that isolates variation in the MSA-level wage time-series due to national level industry trends (Bartik, 1991). To generate our Bartik wages we compute average wages by gender, industry, and MSA using the 2000 census data. We then computed the share of employment by industry and MSA for each month of the sample period using the Bureau of Labor Statistic’s Quarterly Census of Employment and Wages (QCEW). We calculate instrumented wages by assuming that each job pays the estimated wage for the given industry and MSA from 2000. Because QCEW is a census of wages based on unemployment insurance reporting, we get a wage estimate that varies at the MSA month level. As a robustness check we also include data on monthly median rental prices.¹²
- Law enforcement risk. We utilize data from the Law Enforcement Management and Administrative Statistics (LEMAS) databases to determine the number of full time law enforcement employees in each MSA.
- Abuse risk. We use data from the Uniform Crime Reports (UCR) to determine the total annual violent crimes per year in a given MSA as a proxy for abuse risk. We provide a placebo test that the coefficient is not just proxying for overall crime by using the rate of

¹²We thank an anonymous reviewer for suggesting this control.

property crime per capita. These data are both cross-validated with information in the National Incident Based Reporting System (NIBRS).

4 Results

Our core results focus on understanding how risk and labor market opportunities interact. We use straightforward multivariate regression to identify the conditional correlations between various risk factors and pricing.

4.1 Crime, travel time, and pricing

Different service locations expose sex workers to different risks (Spice, 2007; Maticka-Tyndale et al., 1999; Taylor, 2003). Workers in massage parlors are generally considered the safest from abusive customers. Workers who set the service location, i.e. “incall” are the next safest as they can and often do make sure the service takes place in a location with security. Workers offering services in a location of the customer’s choosing, i.e. “outcall,” are the most exposed to risk of abuse by customers or law enforcement stings.

[INSERT FIGURE 4 ABOUT HERE.]

Consistent with workers requiring compensation for risk, there is clear variation in prices by service location. Figure 4 clearly highlights this fact, plotting the distribution of prices for all 2.1M ads with prices which specify a service venue, do not appear to be spam, and are linked to an MSA for which we have the full set of covariates.¹³ Mean and median prices are drastically lower for massage parlor ads than for other service locations. Within other venues both mean and median prices are statistically significantly lower among incall providers than outcall providers, but there is substantial overlap in the price distributions.

Of course, different service locations also require providers to engage in different levels of travel. In particular, providers offering outcall may have to spend more time traveling than those who provider service at a venue of their choosing (incall), and buyers face the opposite costs. We

¹³Although ad content is available through 2016, we are constrained by the availability of MSA-level crime information from the FBI’s uniform crime reports and other covariates from the American Community Survey.

therefore decompose the price premium into two components, physical risk and travel time, by estimating how the relationship between pricing and service venue varies across a proxy for the physical size of the service area, average commute time in the MSA as measured by the American Community Survey.

Specifically, we estimate the following regression at the ad level, dropping all ads by providers who average more than 6 ads per day to remove spam from the sample:

$$\begin{aligned}
 P_{i,j,m} = & \alpha + \beta_1 \text{outcall}_i + \beta_2 \text{both}_{i,j} + \beta_3 \text{unclear}_{i,j} + \beta_4 C_j + \beta_5 (\text{outcall}_{i,j} \times C_j) \\
 & + \beta_6 (\text{both}_{i,j} \times C_j) + \beta_7 (\text{unclear}_{i,j} \times C_j) + \tau_m + \Delta \mathbf{X}_{j,m} + w_j + \epsilon_i. \quad (1)
 \end{aligned}$$

Here ads are indexed by i and each occurs in an MSA j and month m . P_i is the price for an ad and the first three variables in the regression are indicators for whether the ad lists outcall as a location, offers both incall and outcall, or is unclear as to the service venue. β_1 captures the price premium for outcall over incall in an MSA with zero commuting time, β_2 does so for ads offering outcall or incall, and β_3 estimates the price premium for ads which are unclear as to the service venue. β_4 reports how much the price for incall varies as the average commute time in the MSA increases. β_5 through β_7 report how those costs change for other service venues. We include two fixed effects in all regressions, τ_m is a month fixed effect to capture any nationwide secular trends in pricing for sex services in online ads and w_j is a website fixed effect to account for consistent differences in pricing across sites. $\mathbf{X}_{j,m}$ is a vector of MSA/month-level traits such as unemployment rate, number of ads posted, and number of unique providers posting ads, as well as MSA-level variables such as population and racial composition, all of which might be correlated with pricing. We cluster standard errors at the MSA level in all regressions to allow for within locality correlations in the errors when assessing statistical significance.

[INSERT TABLE 3 ABOUT HERE.]

By looking at how much of the price premium for outcall comes from travel costs at different levels of the commute time variable we can assess the relative importance of risk vs. average commute time in pricing. The variation in pricing reflects potential differences in risks and differences

in travel costs for both buyers (men) and sellers (women). To be clear, commute time reported in surveys is a proxy for travel time for the average outcall or incall event, but since there is no reliable way to measure the average travel time for an outcall service event we use it as a proxy. The relationship between the estimated coefficient in table 3 and the value of time to providers (the actual quantity of interest) depends on the ratio of commute time to outcall travel time. If that number is substantially greater than 1, then we understate the value of time and therefore overstate the risk premium providers demand for outcall appointments.¹⁴

As table 3 shows quite clearly the majority of the outcall premium comes from travel time. Column 1 shows the simple differences in costs across service venue, and the price premium is quite clear with outcall services commanding a \$24 per act premium, roughly a 17% premium on the median incall price of \$140. Column 2 adds a control for travel time and various MSA/month and MSA level covariates, showing that commute time is positively correlated with pricing, and that the MSA-level covariates do not affect the price premiums. Column 3 adds in the full set of interaction terms (column 3) to see how incall prices and the outcall premium varies with MSA size. Incall prices are positive but statistically insignificantly correlated with average commute time.¹⁵ However, each additional minute of average commute time predicts a statistically significant increase of \$0.53 in the price of outcall services. At the 50th percentile of commute time, roughly 30 minutes, the outcall premium is \$22.50. The estimates on the interaction terms are quite robust. They change little when we add an MSA fixed-effect (column 4) to account for all time-invariant characteristics of the MSAs or when we include Bartik (1991) instruments for male and female wages in the licit market as controls along with their interaction with commute time (column 5) to account for any correlations driven by secular trends in local labor markets that are not captured by controlling for unemployment rates.¹⁶

¹⁴We thank one of our anonymous reviewers for highlighting this important point.

¹⁵The confidence interval for the relationship between incall pricing and commute time range from -9.4 to 62 at the 99th percentile of commute time.

¹⁶We additionally ensure the results do not mask a non-linear relationship between pricing and travel time by estimating a model in which we interact indicator variables for each decile of commute time with indicators for the different service venues (incall, outcall, both, unclear). This approach allows for the possibility that the outcall premium is flat until very high levels of commute time and other non-linearities. We find that both the outcall premium and the premium for those offering both incall and outcall are generally increasing in commute time. Results available on request.

So what do these prices reveal about the components of pricing? In column 3, our preferred specification because it allows us to directly estimate the role of commute time, we examine the nonlinear effects of commute time on prices. At the 1st percentile of commute time, roughly 20 minutes, the outcall price premium is \$18.3, of which roughly \$11 comes from travel time. At the 50th percentile of commute time, roughly 28 minutes, the outcall premium is \$22.5. And at the 99th percentile of commute time, roughly 38 minutes, the outcall premium is roughly \$28. If we treat the distance-invariant outcall premium in column (3) as the risk component, then the additional risk from performing these services at the buyer's chosen location accounts for roughly 34% of the outcall premium in the median-sized MSA.¹⁷ Thus, travel time appears to be the main driver of outcall pricing but there is another component.

To further assess whether risk is indeed driving price we estimate a series of regressions adding in interactions of service venue with measures of violent crime rates, which we believe proxy for risk to providers and buyers, as well as property crime rates, which are arguably less correlated with risk to providers. As before we show the regression with controls and with MSA fixed-effects and cluster standard errors at the MSA level. If risk is driving a large share of the outcall premium and if incall providers can do more to shield themselves from abuse risk, then we should see that outcall prices are positively correlated with violent crime rates, but incall prices are not. We present the regressions with the interactions of the different types of crime rates with service venues separately (columns 1 and 2 without MSA fixed-effects and columns 4 and 5 with MSA fixed-effects) and jointly (columns 3 and 6) so that we are estimating the correlations with violent crime rates conditional on property crime rates and vice versa.

[INSERT TABLE 4 ABOUT HERE.]

Table 4 shows that prices for outcall services and for services offered in either venue are positively correlated with violent crime rates when controlling for property crime rates (column 3), but the correlation is not robust to controlling for MSA fixed-effects (columns 4 and 6). Conditional on violent crime rates, property crime rates appear to be negatively correlated with pricing for outcall services (columns 3 and 6). The magnitude of the relationship is modest. A one standard deviation

¹⁷The time invariant outcall premium in Column 3 is \$7.58 which is 34% of the \$22.50 outcall premium at the median travel distance.

increase in the number of violent crimes (roughly 310 per 100,000 people per year) predicts a \$4.23 (95% CI of \$0.67 to \$7.8) increase in the outcall premium (using the estimates from column 3). This represents a 0.064 standardized treatment effect and an 2.8% percentage point increase from the mean incall price of \$151/hour. There is no similar positive relationship with property crime rates.

These results are consistent with the incall and outcall markets being segmented. If incall/outcall were one market then we should see prices moving in opposite directions across the markets, that is as conditions favor incall services outcall pricing should drop and vice versa. With respect to commute time, for example, women living in larger areas who do not want to travel (or who live in high-crime areas and fear crime) should pay men to come to them by offering lower prices and they should charge more for travel. That we primarily observe movement in the outcall market across crime rates and city size suggests supply and demand are fairly inelastic with respect to distance in the incall market but not in the outcall market.

Under the assumption that the markets are segmented, then these regressions enable one more decomposition of interest. Sex work is generally considered a distasteful job for which workers require compensation well above the market wages they could earn in other occupations (Rao et al., 2003). The premium that workers in so-called ‘dirty’ jobs receive above what similarly skilled people earn in less unpleasant occupations are known as compensating differentials and can take the form of job amenities (e.g. more time off, flexible hours, etc.) or higher wages (Viscusi, 1993; Lavetti, 2015).¹⁸ Using our data we can break the cost of a session into the travel time component, for which the worker should require no special compensation as it is equivalent to the individuals ‘normal’ work options, and the service time component, for which they would expect special compensation. Assuming that average travel time for an average outcall is similar to the average commute time, then for a median sized city a provider will charge roughly \$18 for 30 minutes spent in the car, implying an hourly wage of \$36/hour, which is small compared to the \$151 mean price for an hour spent with the client in incall services. This is a massive compensating differential, implying either that sex workers are in great demand or that the occupation is quite

¹⁸There is also within-job variation in wages in response to risk. Prostitutes in Mexico, for example, earn a substantial compensating differential for unprotected sex (Rao et al., 2003).

distasteful.¹⁹

As an additional robustness check we include a number of other time-varying controls measured at the MSA-month level in Appendix Table A2.²⁰ Column 1 presents the core estimating equation from Table 3 Column 4 to enable easy comparison. Column 2 adds in controls for the number of ads per capita in that MSA-month to control for the possibility that changes in competition in the ad space differentially affect prices in the outcall vs. incall market. Column 3 adds in linear, quadratic, and cubic terms in the number of providers advertising in the MSA-month for the same reason. Column 4 accounts for the concentration of advertising in the MSI by including a term for the Hirschman-Herfindahl Index (HHI) of the number of providers in the MSA-month. Column 5 includes fixed-effects at the MSA-year level and clusters standard errors at the same level. Column 6 adds controls for the median rental price in the MSA-month to better control for local economic conditions for the subset of 170 MSAs where that variable is available. Column 7 includes all controls for the same subset. Column 8 restricts our core estimating approach to (a) ads that are priced between the 5th and 95th percentile of pricing in all ads and (b) MSA-months that are between the 5th and 95th percentiles in terms of ads per capita. In all cases the core estimate of the interaction term between outcall prices and commute time remains statistically strong and substantively large. In addition we estimate models controlling for rates of prostitution arrests and sex offenses measured at the MSA-month level using UCR data. Although none of the results change with these controls we do not include them in the table because of the significant reporting biases that may be present in using reported crime to measure sexual assault risk and arrest risk.²¹

To account for potentially differential trends across MSA, Appendix Table A3 compares our baseline specification (Column 1) to a model that includes MSA-specific linear time trends (Column 2) and one with quadratic time trends (Column 3). In both cases the core results remain substantively unchanged. Overall the results are quite stable once standard two-way fixed effects are included.

¹⁹As a point of comparison, Lavetti (2015) estimates the compensating differential for fishermen in Alaska (a job carrying an annual risk of death around 1%) of around 4 times higher than other jobs for which individuals in his sample are qualified, which is comparable to our estimate of about 4 times higher wages here.

²⁰We thank our anonymous reviewers for these suggestions.

²¹Results available from authors.

4.2 Controlling for provider characteristics

Our core data lack detailed information on provider traits such as appearance or specific sex acts offered. It is possible, though unlikely, that such traits correlate with MSA size and service location in a way that generates spurious results. As an additional robustness check we re-run the core analysis focusing on prices for providers listed in *The Erotic Review* which provides detailed data on individual providers as described in section 3.2. These data represent the subset of providers advertising online who have chosen to pay for a page on the review site, presumably because doing so enables them to establish a stronger reputation with customers and their average pricing is significantly higher than in the overall sample. If we see a similar outcall premium in this subset where we can control for unobservables in the main data that should provide greater confidence in the core estimates reported in Tables 3 and 4.

Specifically, we estimate the premium associated with offering either outcall or both outcall and incall on the average price charged by a specific service provider:

$$P_{i,j,m} = \alpha + \beta_1 \text{outcall}_i + \beta_2 \text{both}_{i,j} + \beta_3 \text{App}_i + \mathbf{R} \text{Race}_i + \mathbf{D} \text{Desc}_i + \mathbf{A} \text{Act}_i + \beta_4 \text{Perf}_i \\ \tau_m + \gamma_j + \epsilon_{i,j,m}. \quad (2)$$

Here providers are indexed by i and each occurs in an MSA j and month m . P_i is the price for a review and the first two variables in the regression are indicators for whether the provider offers outcall or both incall and outcall services. β_1 captures the price premium for outcall over incall services, while β_2 does so for providers offering outcall or incall services. A series of additional controls are included that account for specific features of the provider. β_3 captures the effect of controls for a provider’s appearance, which is measured on a 1-10 scale. The vector \mathbf{R} identifies the impact of race of the provider, which are measured with a vector of indicator variables of black, asian, hispanic, or other non-white categories. \mathbf{D} captures the price premia associated with a vector of variables describing the provider’s appearance and build, which includes indicators for the provider’s build (e.g. “average” or “thin”), tattoos (e.g. “a few” or “many”), breast appearance (e.g. “average” or “perky”), and breast implants (“yes”, “no”, or “don’t know”). \mathbf{A} identifies the

price premium associated with specific acts performed by the provider, which include indicators for oral sex, oral sex without a condom, anal sex, or multiple orgasms per session. β_4 captures the price premium associated with the provider's performance rating, which is determined on a 1-10 scale. We also include city and month fixed effects, in addition to a provider specific idiosyncratic error term. As before we cluster standard errors at the MSA level.

[INSERT TABLE 5 ABOUT HERE.]

As table 5 indicates, there is a considerable outcall premium. This premium is identifiable regardless of whether the provider offers outcall services exclusively or provides flexibility in their willingness to provide services at the client's location or at the provider's location. Interpreting column (1), we note that relative to provider's that exclusively provide incall services, those that offer outcall or both incall and outcall services receive a price premium of approximately \$57 and \$50, respectively, above the average incall price of \$242.50. In columns (2) - (7) we gradually increase the number of controls that are included in the estimation. In our most saturated model - column (7) - the price premia is considerably smaller than in our most naive specification (approximately half the size), but a sizable and statistically significant effect remains. In effect, after controlling for both observable and a wide range of normally unobservable features of providers, the price premium associated with service providers that are offering higher risk services by traveling to the client's location receive approximately \$20-\$25 more per session, which translates into an approximate 10% price premium above providers that exclusively offer incall services

5 Conclusion

Women selling sex services online appear to engage in rational pricing behavior. The risk associated with traveling to the client is rewarded with a significant price premium that goes beyond the costs associated with commuting. The costs associated with abuse risk borne by direct involvement with the client at a location of their choosing appears to be compensated. Most importantly, the magnitude of the compensating differential workers in this market demand for service time compared to travel time, roughly \$115, is a significant share of the \$151/hour mean price for incall time with a client.

These results imply that many workers in the market would happily shift to other activities given the opportunity and that sex workers advertising online for outcall services have sufficient market power to demand compensation for risks, suggesting the labor supply for this market is fairly inelastic.

More broadly our analysis demonstrates there is great potential for learning about behavior in online markets by combining cutting-edge machine reading technology with established econometric approaches. Importantly, though, if pricing is broadly rational then pricing anomalies should be detectable. Organizations involved in trafficking women for sex work are unlikely to internalize labor market conditions in the same way that voluntary providers do, and so are likely to shift prices in different ways. Such price-based anomaly detection is an important avenue for future research.

Figures

Figure 1: Sample ad from backpage.com.

!!!! SWEET ASHLEY !!!! for a LIMITED TIME Only !!!!! - 27

Posted: Thursday, September 12, 2013 12:03 AM

[Reply](#)

Hello i'm Ashley and I am 100% real and would Love to make your dreams come true

Pic's are 100% real and current, cant wait to meet you.

So call [REDACTED] for the time of your life, Outcalls only

Poster's age: 27

- Location: St. Augustine, St Augustine

- Post ID: 5350879 staugustine

[email to friend](#)

Figure 2: Processing pipeline for sex ads.

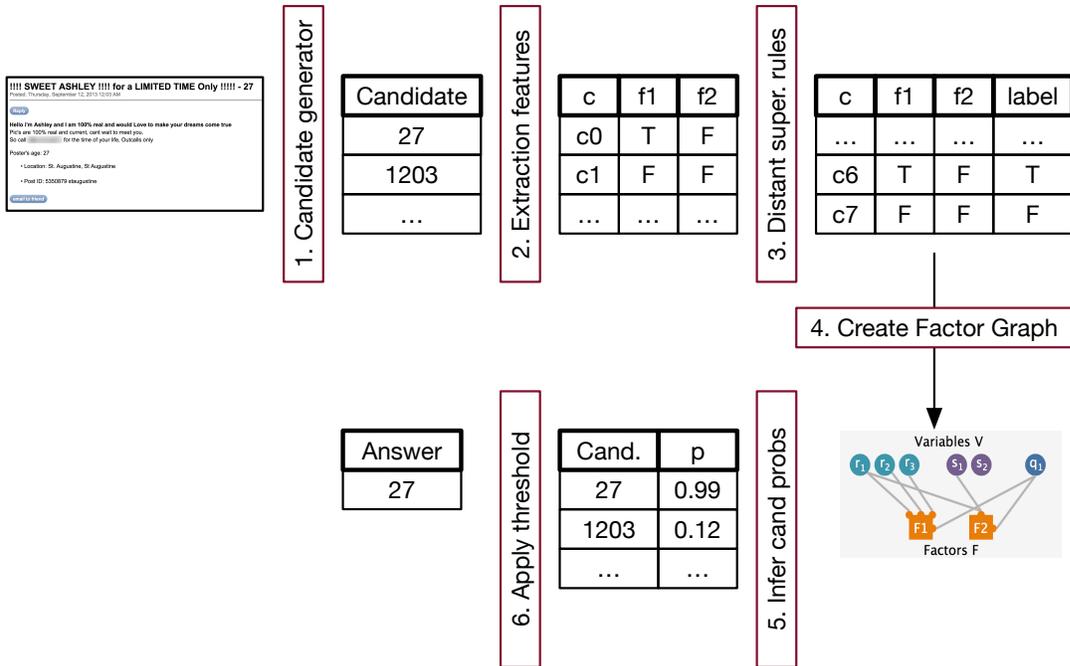


Figure 3: Map of online ads for sex services in the 82 largest MSA throughout the United States. Circles sized by ads per capita. Darker shading indicates higher median prices.

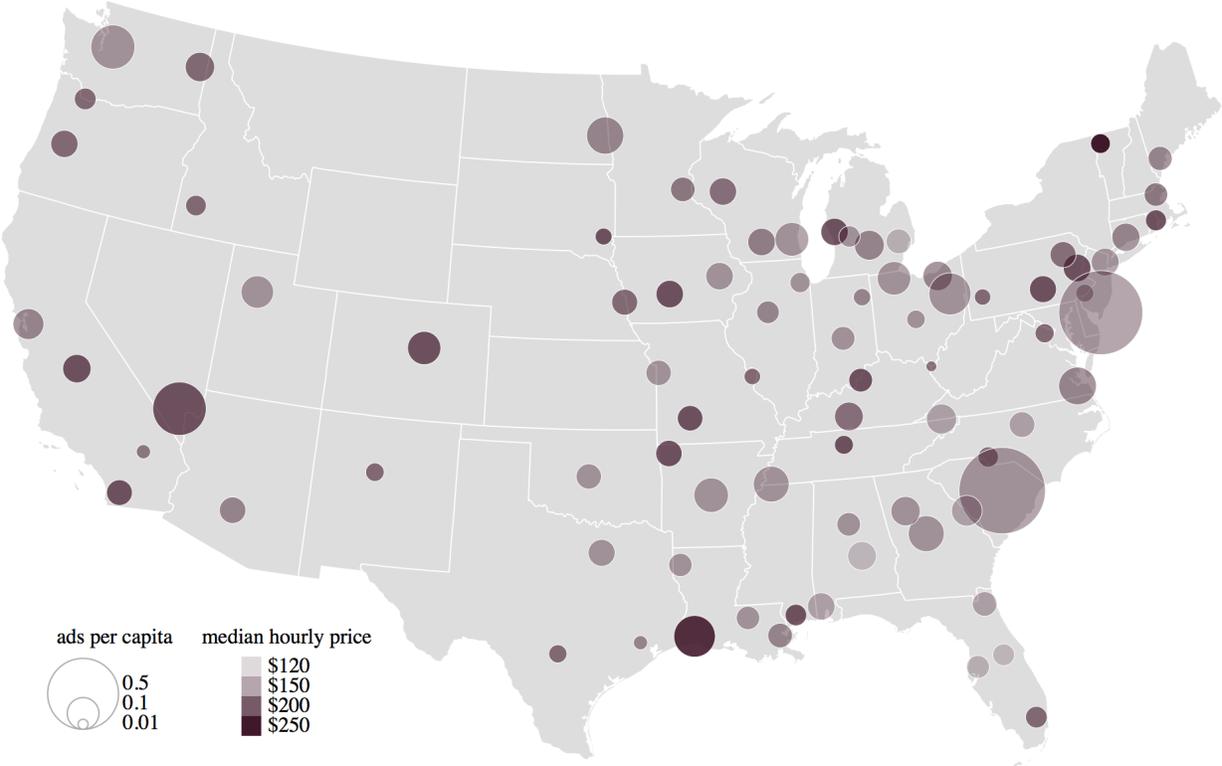


Figure 4: One-hour price distribution by service venue. Prices imputed for ads that do not advertise a one-hour rate.

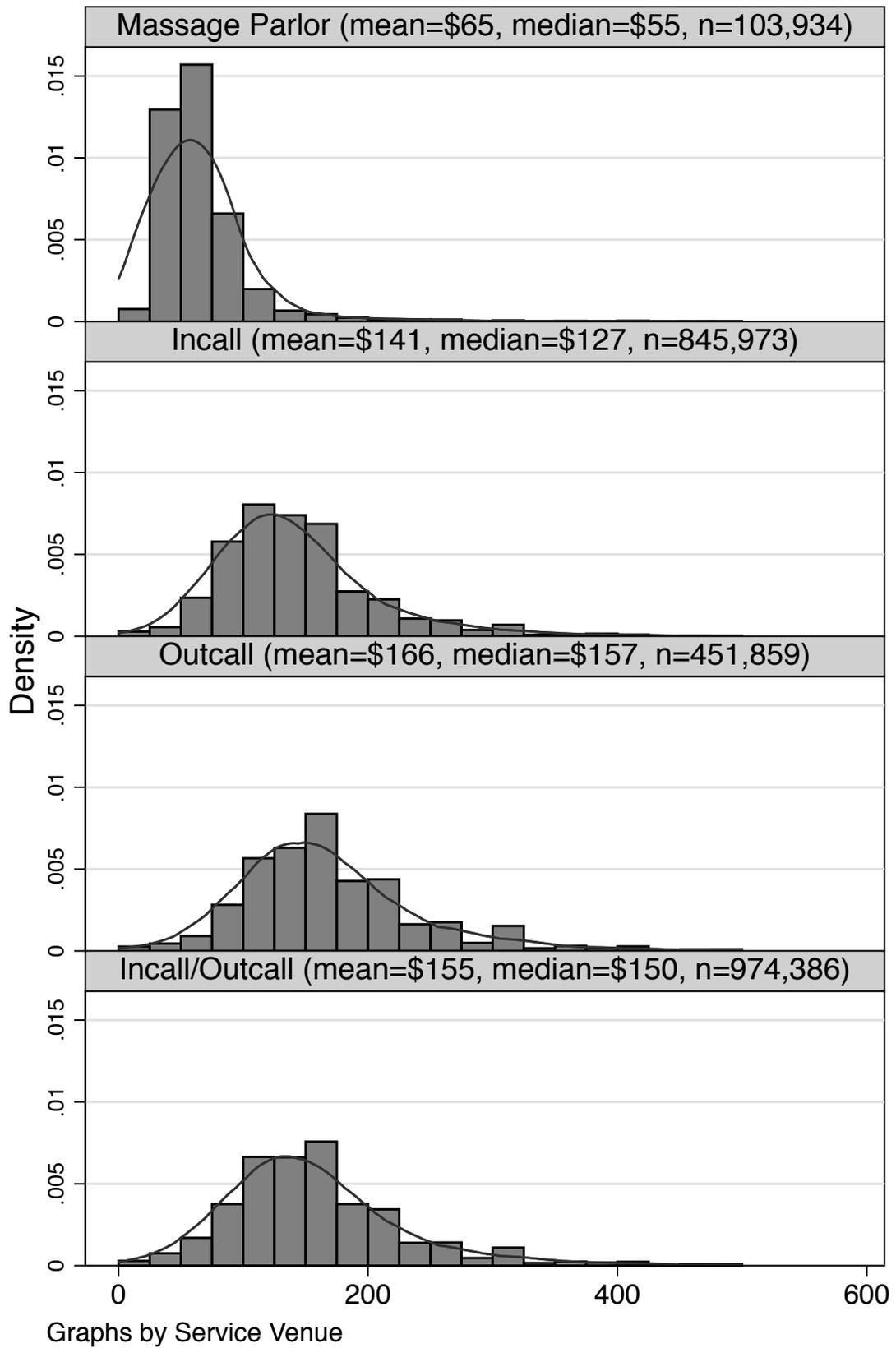


Table 1: Summary Statistics - Ad Data and MSAs

Variable	Observations	Mean	Std. Dev.	Min	Max
Panel A: Service Price and Type					
Price per hour	2,732,414	149.63	75.77	10	1100
Incall	2,732,414	.33		0	1
Outcall	2,732,414	.17		0	1
Both	2,732,414	.38		0	1
Unclear	2,732,414	.076		0	1
Massage parlor	2,732,414	.042		0	1
Panel B: Service Price and Type - Estimation Sample					
Price per hour	2,098,241	151.7	67.7	10	500
Incall	2,098,241	.34		0	1
Outcall	2,098,241	.18		0	1
Both	2,098,241	.39		0	1
Unclear	2,098,241	.08		0	1
Massage parlor	2,098,241	0		0	0
Panel C: MSA/month controls					
Unemployment (BLS)	7,226	.095	0.02	0.04	0.18
Female Bartik wages	7,226	14.0	2.20	9.60	23
Male Bartik wages	7,226	22	3.1	16	33
Ad Count	7,226	320	741	1	12541
Unique providers	7,226	145	343	1	7333
Panel D: MSA level controls					
Average commute time	187	25	3.4	18	38
Population (millions)	187	.78	1.4	.07	13
Share White	187	.78	.11	.48	.95
Violent Crimes per 100,000	187	569	341	0	1668
Property Crimes per 10,000	187	437	232	0	899

Table 2: Summary Statistics - Reviews

Variable	Observations	Mean	Std. Dev.	Min	Max
Panel A: Service Price and Type					
Price per hour	1,097,527	276	119	50	1,000
Incall	1,097,527	.35		0	1
Outcall	1,097,527	.09		0	1
Both	1,097,527	.56		0	1
Panel B: Provider Traits					
Appearance (1-10)	1,097,527	7.5	1.3	1	10
Photo accurate	1,097,527	.87		0	1
Age	1,097,527	29.1	7.1	18	51
Height (inches)	1,097,527	65.0	2.9	56	78
Black	1,101,788	.08		0	1
Asian	1,101,788	.16		0	1
Hispanic	1,101,788	.11		0	1
White	1,101,788	.56		0	1
Panel C: Select Review Elements					
Independent	229,068	.52		0	1
Months first to last review	229,068	7.9	17.5	1	205
Number of unique reviews per provider	229,068	4.8	13.4	1	853
Providers per city	91	2,517	3,991	1	19,615

Notes: Data from www.theeroticreview.com, scraped in March 2016. Price per hour calculated for all reviewed providers. Independent refers to providers listed as being independent. Of remainder 70.4% are listed as working with an agency, 26.5% as “don’t know”, and 3.1% as working with a driver. Independent providers had 6.1 reviews on average compared to 3.4 on average for others. Sample restricted to reviews with realistic prices ($\leq \$50$), non-missing appearance ($n = 2,108$ reviews are missing appearance rating).

Table 3: Commute time and prices in the online market for sex

	(1)	(2)	(3)	(4)	(5)
Outcome	Price	Price	Price	Price	Price
Outcall	23.601*** (1.160)	23.022*** (1.095)	7.758 (8.361)	2.826 (7.956)	2.808 (7.937)
Both	13.640*** (0.915)	13.306*** (0.904)	3.784 (5.900)	1.567 (5.894)	1.565 (5.907)
Unclear	5.738*** (1.019)	5.705*** (0.965)	0.502 (4.988)	1.114 (4.729)	1.155 (4.721)
Avg. Commute		0.881** (0.389)	0.636 (0.387)		
Outcall \times Commute			0.530* (0.297)	0.645** (0.286)	0.646** (0.285)
Both \times Commute			0.332 (0.205)	0.381* (0.207)	0.381* (0.207)
Unclear \times Commute			0.180 (0.174)	0.144 (0.165)	0.142 (0.165)
Outcall premium: 1st pctile commute (20 min)			18.4*** (2.6)	15.8*** (2.4)	15.8*** (2.4)
Median commute (28 min)			22.5*** (1.1)	20.8*** (1.0)	20.8*** (1.0)
99th pctile commute (38 min)			28.0*** (3.3)	27.5*** (3.2)	27.4*** (3.2)
R-Squared	0.060	0.069	0.070	0.097	0.097
Observations	2,098,241	2,098,241	2,098,241	2,098,241	2,098,241
Clusters	182	182	182	182	182
Additional MSA controls		Y	Y	Y	Y
MSA Fixed Effects				Y	Y
Bartik Wages					Y

Notes: Prices calculation include only providers who post 200 or fewer times to omit spam. All regressions include month and website fixed effects. Additional controls include number of ads posted in the MSA/month, number of unique providers posting in MSA/month, population, unemployment rate, and fraction of the MSA population that is white. Estimates significant at the 0.05 (0.10, 0.01) level are marked with ** (*, ***). Standard errors are clustered at the MSA level.

Table 4: Commute time and prices in the online market for sex

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	Price	Price	Price	Price	Price	Price
Outcall	6.941 (8.238)	13.530 (9.123)	16.868* (9.585)	4.452 (7.991)	9.493 (8.820)	10.872 (9.091)
Both	1.273 (5.837)	4.573 (6.845)	7.470 (6.683)	0.765 (5.927)	3.382 (6.955)	5.020 (7.066)
Unclear	2.490 (5.147)	5.503 (6.294)	5.629 (6.464)	3.494 (4.841)	6.347 (6.093)	6.176 (6.224)
Violent crimes per 100,000	-0.006 (0.005)		-0.002 (0.008)			
Outcall \times Violent crimes per 100,000	0.001 (0.005)		0.014** (0.006)	-0.003 (0.004)		0.005 (0.005)
Both \times Violent crimes per 100,000	0.004 (0.003)		0.013** (0.005)	0.001 (0.003)		0.007 (0.005)
Unclear \times Violent crimes per 100,000	-0.003 (0.004)		0.001 (0.005)	-0.004 (0.004)		-0.000 (0.005)
Property crimes 10,000		-0.013** (0.006)	-0.011 (0.012)			
Outcall \times Property crimes 10,000		-0.008 (0.005)	-0.024*** (0.009)		-0.009* (0.005)	-0.016* (0.008)
Both \times Property crimes 10,000		-0.001 (0.004)	-0.016** (0.007)		-0.003 (0.004)	-0.011 (0.007)
Unclear \times Property crimes per 10,000		-0.006 (0.005)	-0.007 (0.008)		-0.007 (0.005)	-0.007 (0.007)
R-Squared	0.070	0.071	0.071	0.097	0.097	0.097
Observations	2,098,241	2,098,241	2,098,241	2,098,241	2,098,241	2,098,241
Clusters	182	182	182	182	182	182
Additional controls	Y	Y	Y	Y	Y	Y
MSA Fixed Effects				Y	Y	Y

Notes: Prices calculation include only providers who post 200 or fewer times to omit spam. All regressions include controls for commute time and its interactions with service venues as well as month and website fixed effects. Additional controls include number of ads posted in the MSA/month, number of unique providers posting, population, unemployment rate, and fraction of the MSA population that is white. Estimates significant at the 0.05 (0.10, 0.01) level are marked with ** (*, ***). Standard errors are clustered at the MSA level.

Table 5: Provider traits and prices in the online market for sex

	Incall Price (mean=242.5; sd=4.7)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcall Only	56.773*** (16.246)	63.013*** (9.984)	44.490*** (7.682)	38.913*** (6.415)	39.276*** (6.390)	34.660*** (6.380)	25.821*** (5.684)
Outcall and Incall	49.821*** (5.584)	54.368*** (6.467)	35.361*** (4.044)	30.559*** (3.043)	31.082*** (3.066)	24.240*** (2.710)	20.577*** (2.649)
City and Month FE		X	X	X	X	X	X
Appearance Controls			X	X	X	X	X
Racial Category				X	X	X	X
Description					X	X	X
Acts Offered						X	X
Performance Rating							X
R-Squared	0.041	0.132	0.366	0.382	0.387	0.403	0.439
Observations	1,101,747	1,101,747	1,101,747	1,101,747	1,101,747	1,101,747	1,101,747
Clusters	91	91	91	91	91	91	91

Notes: price per hour is average across reviews. Appearance controls include cubic terms in average subject appearance rating (1-10 scale) for provider, height and age (results similar if use categorical variable for age or height, nested model tests show no significant increase in in-sample explanatory power from including higher-order terms in age and height). Racial category controls include dummy variables for providers reported as being black, asian, hispanic, or other non-white categories. Description controls include dummy variables for various descriptions of build (e.g. “average” or “thin”, etc.), tattoos (e.g. “a few” or “many”, etc.), breast appearance (e.g. “average” or “perky”, etc.), and whether the provider has breast implants (“yes”, “no”, or “don’t know”). Sex act controls include dummy variables for whether the reviewer reports the provider offering oral sex, oral sex without condom, anal sex, or multiple orgasms per appointment. Performance rating includes cubic term in average subjective performance rating (1-10 scale) for the provider. Estimates significant at the 0.05 (0.10, 0.01) level are marked with ** (*, ***). Standard errors are clustered at the city level.

References

- Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Number wbsle in “Books from Upjohn Press” W.E. Upjohn Institute for Employment Research.
URL: <https://ideas.repec.org/b/upj/ubooks/wbsle.html>
- Bass, Allison. 2015. *Getting Screwed: Sex Workers and the Law*. University Press of New England.
- Blanchard, Olivier Jean and Lawrence F. Katz. 1992. “Regional Evolutions.” *Brookings Papers on Economic Activity* (1):1–75.
- Callaway, Ewen. 2015. “Computers Read the Fossil Record.” *Nature* .
- Cunningham, Scott, Gregory DeAngelo and John Tripp. 2018. “Craiglist’s Effect on Violence Against Women.” *Working Paper* .
- Cunningham, Scott and Manisha Shah. 2014. Decriminalizing Indoor Prostitution: Implications for Sexual Violence and Public Health. Working Paper 20281 National Bureau of Economic Research.
URL: <http://www.nber.org/papers/w20281>
- Cunningham, Scott and Todd D. Kendall. 2011*a*. “Men in Transit and Prostitution: Using Political Conventions as A Natural Experiment.” *The B.E. Journal of Economic Analysis & Policy* 11(1).
- Cunningham, Scott and Todd D. Kendall. 2011*b*. “Prostitution 2.0: The changing face of sex work.” *Journal of Urban Economics* 69(3):273–287.
- Cunningham, Scott and Todd D. Kendall. 2011*c*. *Prostitution, technology and the law: new data and directions*. Edward Elgar Publishing Limited chapter Research Handbook in the Law and Economics of the Family.
- Cunningham, Scott and Todd D. Kendall. 2016. Examining the Role of Client Reviews and Reputation Within Online Prostitution. In *The Oxford Handbook of the Economics of Prostitution*, ed. Scott Cunningham and Manisha Shah. Oxford University Press chapter 2.
- Dank, Meredith, Bilal Khan, P. Mitchell Downey, Cybele Kotonias, Debbie Mayer, Colleen Owens, Laura Pacifici and Lilly Yu. 2014. *Estimating the Size and Structure of the Underground Commercial Sex Economy in Eight Major US Cities*. Urban Institute.
- Delap, Josie. 2014. “More bang for your buck.”
- Edlund, Lena, Joseph Engelberg and Christopher A. Parsons. 2009. The Wages of Sin. Economics Discussion Paper 0809-16 Columbia University.
- Gertler, Paul, Manisha Shah and Stefano M. Bertozzi. 2005. “Risky Business: The Market for Unprotected Commercial Sex.” *Journal of Political Economy* 113(3):518–550.
- Gould, Eric D., Bruce A. Weinberg and David B. Mustard. 2002. “Crime Rates and Local Labor Market Opportunities in the United States: 1979-1997.” *The Review of Economics and Statistics* 84(1):45–61.
- Hubbard, Phil and Teela Sanders. 2003. “Making Space for Sex Work: Female Street Prostitution and the Production of Urban Space.” *International Journal of Urban and Regional Research* .

- Hughes, Donna M. 2002. “The Use of New Communications and Information Technologies for sexual Exploitation of Women and Children.” *Hastings Women’s Law Journal* 13:129–148.
- Kolker, Robert. 2013. “The New Prostitutes.” *The New York Times* .
- Latonero, Mark, Genet Berhane, Ashley Hernandez, Tala Mohebi and Lauren Movius. 2011. Human Trafficking Online: The Role of Social Networking Sites and Online Classifieds. Report USC Annenberg Center on Communication Leadership & Policy.
- Lavetti, Kurt. 2015. Estimating Preferences in Hedonic Wage Models: Lessons from the *Deadliest Catch*. Working paper Ohio State University.
- Logan, Trevor D. and Manisha Shah. 2013. “Face Value: Information and Signaling in an Illegal Market.” *Southern Economic Journal* 79(3):529–564.
- Maticka-Tyndale, Eleanor, Jacqueline Lewis, Jocalyn P Clark, Jennifer Zubick and Shelley Young. 1999. “Social and cultural vulnerability to sexually transmitted infection: the work of exotic dancers.” *Canadian journal of public health* 90(1):19.
- Mitchell, Kimberly J., David Finkelhor, Lisa M. Jones and Janis Wolak. 2010. “Use of Social Networking Sites in Online Sex Crime Against Minors: An Examination of National Incidence and Mean of Utilization.” *Journal of Adolescent Health* 47(2):183–90.
- Moffatt, Peter G. and Simon A. Peters. 2004. “Pricing Personal Services: An Empirical Study of Earnings in the UK Prostitution Industry.” *Scottish Journal of Political Economy* 51(5):675–690.
- More Bang for your Buck*. 2014.
- Peters, Shanan E., Ce Zhang, Miron Livny and Christopher Ré. 2014. “A Machine Reading System for Assembling Synthetic Paleontological Databases.” *PLoS ONE* .
- Potterat, John J., Devon D. Brewer, Stephen Q. Muth, Richard B. Rothenberg, Donald E. Woodhouse, John B. Muth, Heather K. Stites and Stuart Brody. 2004. “Mortality in a Long-term Open Cohort of Prostitute Women.” *American Journal of Epidemiology* 159(8):778–785.
- Rao, Vijayendra, Indrani Guptab, Michael Lokshina and Smarajit Janac. 2003. “Sex workers and the cost of safe sex: the compensating differential for condom use among Calcutta prostitutes.” *Journal of Development Economics* 71(2):585–603.
- Recht, Benjamin, Christopher Re, Stephen J. Wright and Feng Niu. 2011. Hogwild: A Lock-Free Approach to Parallelizing Stochastic Gradient Descent. In *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain*. pp. 693–701.
- Roe-Sepowitz, Dominique, Kristine Hickle, James Gallagher and Jessica Smith. 2016. “Invisible Offenders: A Study Estimating Online Sex Customers.” *Journal of Human Trafficking* pp. 272–280.
- Shin, Jaeho, Sen Wu, Feiran Wang, Christopher De Sa, Ce Zhang and Christopher Ré. 2015. “Incremental Knowledge Base Construction Using DeepDive.” *PVLDB* 8(11):1310–1321.
- Spice, William. 2007. “Management of sex workers and other high-risk groups.” *Occupational medicine* 57(5):322–328.

- Taylor, D. 2003. *Sex for Sale: New Challenges and New Dangers for Women Working on and off the Streets*. Mainliners.
- Viscusi, W. Kip. 1993. “The Value of Risks to Life and Health.” *Journal of Economic Literature* 31:1912–1946.
- Weitzer, Ronald. 1999. “Prostitution Control in America: Rethinking Public Policy.” *Crime, Law and Social Change* 32(1).
- Zhang, Ce. 2015. DeepDive: A Data Management System for Automatic Knowledge Base Construction PhD thesis University of Wisconsin-Madison.
- Zhang, Ce and Christopher Re. 2014. “DimmWitted: A Study of Main-Memory Statistical Analytics.” *PVLDB* 7(12):1283–1294.

Appendix

This appendix provides additional supporting materials.

Extraction Performance

Table A1 provides summary statistics on the extraction of various data fields from the ad text.

Table A1: Extraction counts and precision/recall statistics

Number of raw advertisements	4,491,779
Number of unique providers	1,666,284
Number of unique websites	19
Number of unique locations	438
Email precision / recall	100/100
Phone precision / recall	100/94
Service venue precision / recall	89/97
Price precision / recall	93/82
Location precision / recall	97/88

Notes: E-mail and phone number extractions based on regular expression matching, not statistical inference. All other outcomes derived using DEEPDIVE.

Incall vs. Outcall Price Distribution

Figure A1: Distribution of average prices for incall and outcall by MSA.

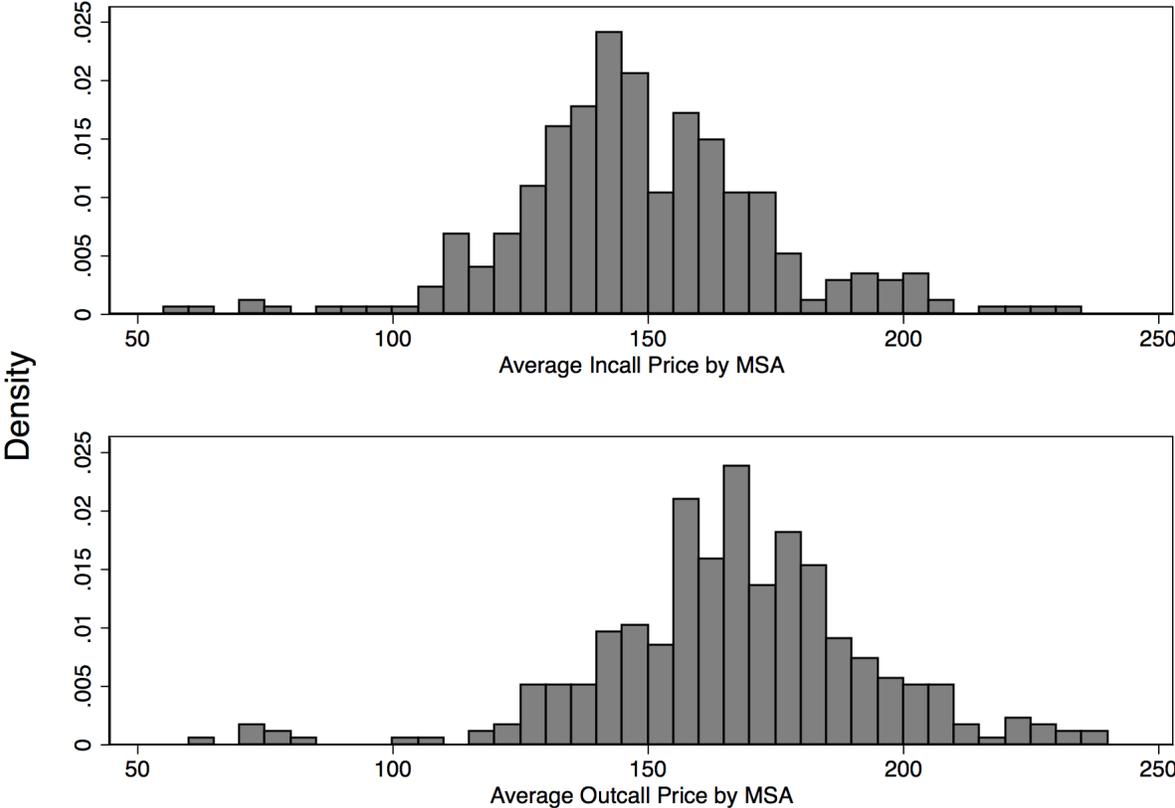
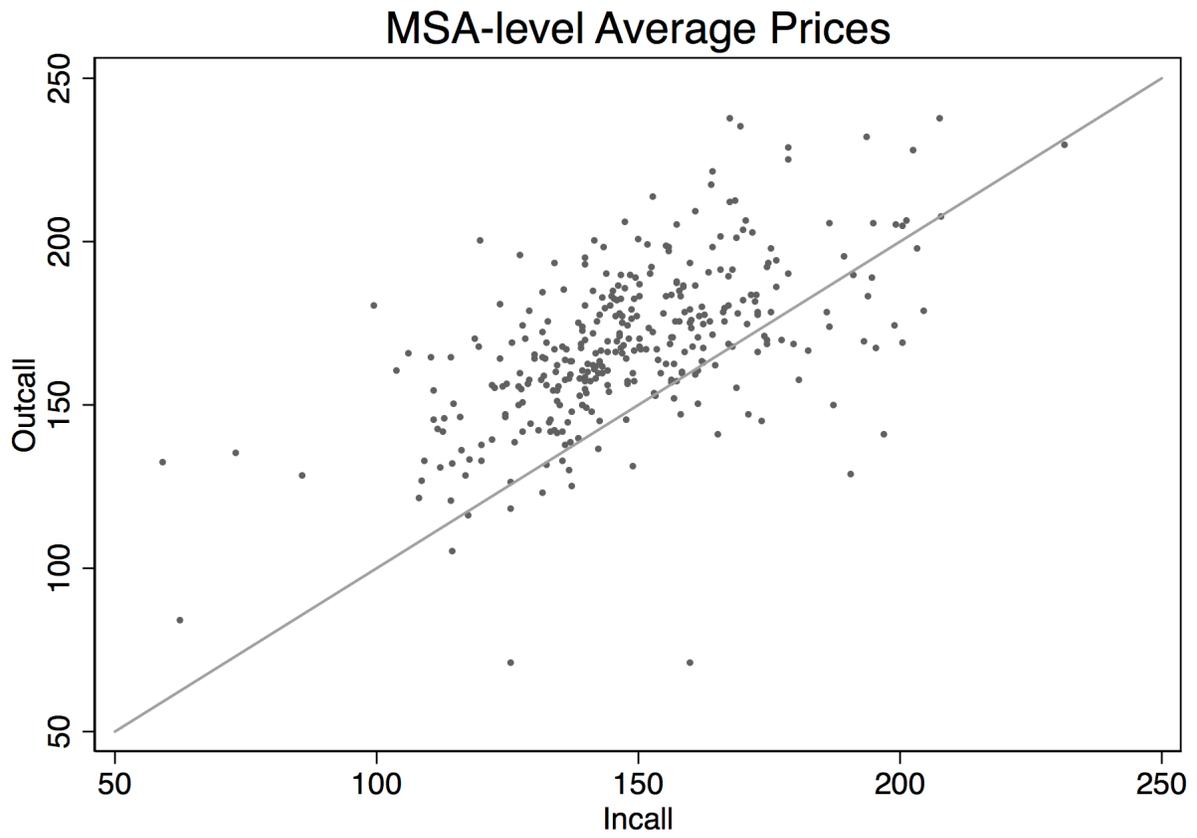


Figure A2: Mean outcall price vs. mean incall price by MSA.



Robustness Checks

Table A2: Commute time and prices with additional controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome	Price	Price	Price	Price	Price	Price	Price	Price
Outcall	2.826 (7.956)	2.964 (7.957)	2.756 (7.970)	2.850 (7.961)	2.867 (5.208)	-5.584 (9.784)	-5.377 (9.768)	-2.156 (3.954)
Both	1.567 (5.894)	1.743 (5.876)	1.534 (5.926)	1.586 (5.901)	1.223 (4.061)	7.795 (7.725)	7.987 (7.647)	2.652 (3.683)
Unclear	1.114 (4.729)	0.994 (4.737)	0.979 (4.756)	1.123 (4.730)	1.162 (4.905)	4.388 (5.310)	4.352 (5.307)	5.504 (5.934)
Outcall × Commute	0.645** (0.286)	0.640** (0.286)	0.648** (0.286)	0.644** (0.286)	0.642*** (0.188)	0.983*** (0.351)	0.975*** (0.350)	0.626*** (0.133)
Both × Commute	0.381* (0.207)	0.375* (0.206)	0.382* (0.208)	0.381* (0.207)	0.391*** (0.144)	0.164 (0.280)	0.158 (0.278)	0.234* (0.128)
Unclear × Commute	0.144 (0.165)	0.148 (0.165)	0.149 (0.166)	0.143 (0.165)	0.142 (0.178)	0.025 (0.193)	0.025 (0.193)	-0.029 (0.216)
R-Squared	0.097	0.097	0.097	0.097	0.104	0.104	0.104	0.104
Observations	2,098,241	2,098,241	2,098,241	2,098,241	2,098,241	1,672,591	1,672,591	1,528,725
Clusters	182	182	182	182	807	170	170	182
Baseline model (Table 3 Column 4)	Y	Y	Y	Y	Y	Y	Y	Y
Additional MSA controls	Y	Y	Y	Y	Y	Y	Y	Y
MSA fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Ads per capita		Y					Y	
Cubic term in provider count			Y				Y	
HHI of providers				Y			Y	
MSA × Year FE					Y		Y	
Median rental prices						Y	Y	
Dropping outliers								Y

Notes: Prices calculation include only providers who post less than 6 ads per day to omit spam. All regressions include month and website fixed effects. Additional controls include number of ads posted in the MSA/month, number of unique providers posting in MSA/month, population, unemployment rate, and fraction of the MSA population that is white. Estimates significant at the 0.05 (0.10, 0.01) level are marked with ** (*, ***). Standard errors are clustered at the MSA level.

Table A3: Commute time and prices with MSA-specific time trends

	(1)	(2)	(3)
Outcome	Price	Price	Price
Outcall	2.826 (7.956)	2.582 (7.950)	2.467 (7.904)
Both	1.567 (5.894)	1.426 (5.853)	1.015 (5.834)
Unclear	1.114 (4.729)	1.117 (4.786)	1.131 (4.848)
Outcall \times Commute	0.645** (0.286)	0.653** (0.286)	0.655** (0.284)
Both \times Commute	0.381* (0.207)	0.386* (0.205)	0.398* (0.204)
Unclear \times Commute	0.144 (0.165)	0.146 (0.167)	0.145 (0.169)
R-Squared	0.097	0.099	0.101
Observations	2,098,241	2,098,241	2,098,241
Clusters	182	182	182
Baseline model (Table 3 Column 4)	Y	Y	Y
MSA-specific linear trends		Y	Y
MSA-specific quadratic trends			Y

Notes: Prices calculation include only providers who post less than 6 ads per day to omit spam. All regressions include month and website fixed effects. Additional controls include number of ads posted in the MSA/month, number of unique providers posting in MSA/month, population, unemployment rate, and fraction of the MSA population that is white. Estimates significant at the 0.05 (0.10, 0.01) level are marked with ** (*, ***). Standard errors are clustered at the MSA level.