AI, Skill, and Productivity

The Case of Taxi Drivers

Kyogo Kanazawa[*] Daiji Kawaguchi[*†]
Hitoshi Shigeoka[*§] Yasutora Watanabe[*]

[*] University of Tokyo [†] RIETI and IZA [§] Simon Fraser University and NBER

December 8, 2021

Motivation: AI and its impact

- AI has a potential to drastically reshape employment.
- However, the impact of AI on employment could be fundamentally different from past technologies.
  - IT and robots replaced routine and manual tasks
  - These technologies are skill augmenting and inequality enhancing
  - On the other hand, AI will replace non-routine cognitive tasks
- Studies (e.g., Webb, 2020) show AI affects *high-skilled* occupations
  - AI might reduce the (wage) inequality among workers

Motivation: AI and its impact (cont.)

- All existing studies examine the impact of AI *across* occupations.
  - They use variation in exposure of each occupation to AI at task level
  - They implicitly assume that workers *within* the occupation are *uniformly* affected by AI
- The impact of AI can be more complex and nuanced
- Even *within* an occupation, there is a substantial heterogeneity of skill for the task that can be replaced by AI.
  - Examining the heterogeneous impact of AI by skill *within* an occupation provides a deeper understanding of AI's complex impact.
- Micro-level evidence/data at this granularity is scant (except for Grennan and Michaely 2020)

Objective

- We study the impact of AI on productivity across workers with different skill levels in the context of taxi drivers.
- Taxi drivers provide an ideal set-up
  - Drivers work independently, thus individual productivity can be easily measured
  - Confounding factors for of productivity gain are limited: use the same capital (taxi), output and input prices are almost the same for all drivers

Our setup: AI Navi app

- The particular AI application we study is called "AI Navi"
  - AI Navi helps drivers search for customers when a taxi is vacant.
  - When it is turned on, AI Navi suggests the routes based on the predicted demand to maximize the probability a taxi will catch customers given the location of the taxi
  - AI Navi's demand forecast is trained by past taxi driving data in the area
- This type of AI, which increases accuracy of *prediction* using machine learning, is widely used in business settings

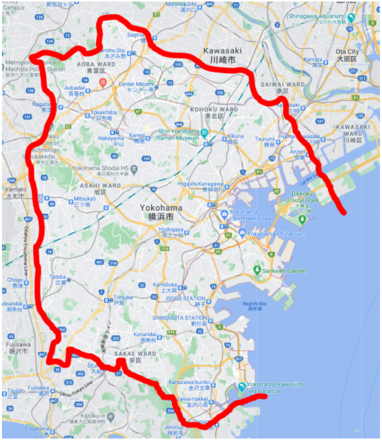
© Zenrin © Mapbox

- The figure displays the snapshot of AI Navi when it is turned on
- AI Navi shows the suggested routes in green with a red arrow given a taxi's current location. See video.
- Red dots indicate locations with potentially high demand

Impact of AI and its relationship with skill

- AI Navi may improve productivity by reducing the search time
  - Taxi drivers spends more than 80% (!) of time searching for customers
  - Thus, we measure productivity by length of search time
- To the extent that demand-forecasting skill is an important component of taxi drivers' skill set, the impact of AI may differ by the drivers' skill
  - Demand forecasting is a non-routine process for which information processing is crucial
- Whether the AI *complements* or *substitutes* the demand forecasting skill of drivers is an empirical question
  - If AI Navi and driver skill are *complement*, AI Navi's impact would be larger for *more* skilled workers
  - If *substitute*, the impact would be larger for *less* skilled drivers

Background: Yokohama city



- Yokohama city has population of 3.75 million
- It has 18 wards with area of 435km^2 (7 times the area of Manhattan)

Background: Taxi drivers in Yokohama

- 4,691 taxis with 8,842 drivers in Yokohama
- Average age of drivers is 61.2
- Yokohama taxi drivers are allowed to drop-off anywhere, but not allowed to pick up outside Yokohama area (Yokohama city and three other nearby cities: Kawasaki, Yokosuka, and Miura).
- Drivers have incentives to increase sales
  - Drivers are paid by the fixed percentage of the fares they collect (usually between 45% and 60%)
  - Drivers do not incur any variable costs (including gas)

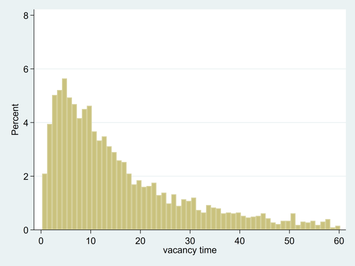
Data

- The data are detailed driving data from the tech company that developed AI Navi
  - To gather field data, AI Navi was provided to 557 taxi drivers (5.9% of Yokohama drivers) *for free* between Dec.3 and Dec 31, 2019
  - Drivers have discretion on usage of AI Navi
  - Additional driving data for Oct and Nov 2019 (pre-trial period)
- Unit of observation is a vacant cruise
  - Vacant cruise time is defined as time between dropping-off the last customer and picking-up a new customer on the streets

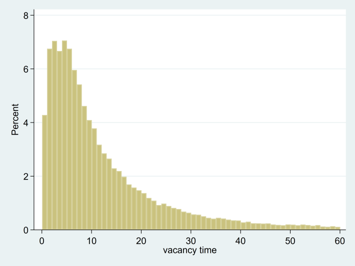
Summary statistics

- Original data consists of 67,111 vacant cruises
- We exclude the followings vacant cruises
  1. 927 vacant cruises of drivers whose Oct and Nov data are not available
  2. 2,758 vacant cruises whose cruise times are more than 60 minutes to exclude outliers
- Data for estimation are 63,426 vacant cruises
  - AI Navi is turned on in 3,242 (5.1%) and off in 60,184 (94.9%)
  - Out of 524 drivers, 201 (38.4%) drivers used AI Navi at least once

Histogram of Vacant Cruise Time: AI is On/Off



(a) AI is turned on



(b) AI is turned off

- The mean(median) vacant cruise time is:
  - (a) AI on 15.8(11.7) minutes
  - (b) AI off 11.9(8.05) minutes
- This suggests selection: drivers turn on AI when it is difficult to find customers

Empirical Strategy

- We compare the vacant cruise time between AI is on and off *within* the same driver in the *similar* demand condition by including 18 ward FEs, and 696 date-hour FEs (=29 days$\times$24 hours/day).
- We estimate Weibull hazard model of finding customers
- Probability that vacant cruise time T become larger than t is denoted as a survival function $S(t) = \Pr(T > t)$ where

$$\begin{aligned} S_{ijhs}(t) &= \exp(-\lambda_{ijhs}(t) \cdot t^p) \\ \lambda_{ijh}(t) &= \exp\{-p(\alpha \cdot \text{AI Navi usage Dummy}_{ijhs,t} + \text{driver FE}_i \\ &\quad + \text{ward FE}_j + \text{date-hour FE}_h)\} \end{aligned}$$

i:driver, j:ward, h:date-hour, s:vacant cruise, t:time of vacant cruise

Empirical Strategy

- The model can be interpreted as

$$\begin{aligned}&\log(\text{vacant cruise time}_{ijhs})\\=&\alpha\cdot\text{AI Navi usage Dummy}_{ijhs,t}\\&+\text{driver FE}_i+\text{ward FE}_j+\text{date-hour FE}_h+\epsilon_{ijhs}\end{aligned}$$

- ϵ follows an extreme-value distribution (Wooldridge 2010, p. 998)
- α corresponds to percentage change in vacant cruise time ($\alpha < 0$).

Identification

- Identification is from within-driver, within-ward, within-date-hour variation in AI Navi usage and the hazard rate of finding customer.
- We assume that AI Navi usage is quasi-random after controlling for driver FE, ward FE, and date-hour FE (we check this later).
- This may be likely because all drivers are new to AI Navi, and they may experiment to turn it on and off.
- To the extent that drivers are likely to turn on AI Navi when the demand is *low* (and finding customers is difficult), the bias goes *against* our finding that AI Navi reduces the search time.
    - Our estimates is a *lower* bound in magnitude for the effect of AI.

Skill and Vacancy Indices

- To understand the effects of driver skill and demand condition, we construct the following two indices.
- Here, we use the vacant cruise data for Oct and Nov *before* our sample period.
- Driver skill index
  - We regress above-mentioned hazard model of finding customer onto driver FE, ward FE, and date-hour FE.
  - Then, standardize the estimated driver FE. The *higher* the driver skill index, the *more* skilled driver.
- Vacancy index
  - Similarly, we construct vacancy index that is ward-day-hour FE (e.g., 10 pm on Wednesday at Ward 1).

Credibility of underlying assumption

Logit of AI Navi usage	(1) Full	(2) Full	(3) Navi users	(4) Navi users	(5) Navi users
skill index	-0.196	-0.190	-0.190	-0.175	
	(0.137)	(0.143)	(0.116)	(0.119)	
vacancy index	0.098***	-0.039	0.091***	-0.034	-0.033
	(0.030)	(0.029)	(0.031)	(0.031)	(0.046)
driver FE					✓
ward FE		✓		✓	✓
date-hour FE		✓		✓	✓
N	63,208	56,416	28,804	25,930	25,930

SEs are in parenthesis clustered on drivers. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

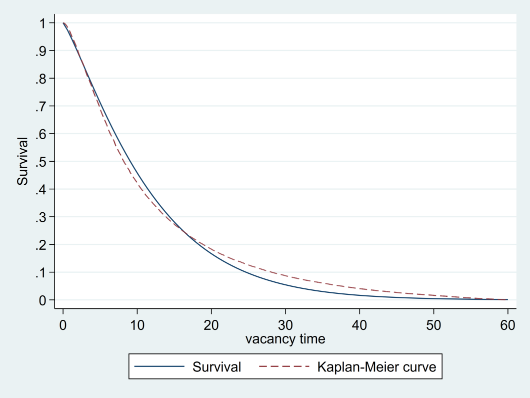
- AI Navi is more likely to be used when demand is low (column 1).
- With ward and date-hour FEs, vacancy index does not matter any more (column 2), suggesting that the demand condition can be well-controlled by ward and date-time FEs.
- With these FEs, usage of AI Navi can be views as good as random

Effect on overall productivity

	(1) Full	(2) Navi users	(3) Navi users $0.1 \leq PS \leq 0.9$
Navi usage	-0.055**	-0.049**	-0.069***
	(0.022)	(0.022)	(0.026)
$\log(p)$	0.181***	0.188***	0.248***
	(0.005)	(0.006)	(0.011)
driver FE	✓	✓	✓
ward FE	✓	✓	✓
date-hour FE	✓	✓	✓
N	63,426	28,909	6,567

- AI Navi usage shorten the time to find customers by about 5%.
- $log(p) > 0$ implies hazard increasing with vacant time
- The result is robust to
  - including AI Navi non-users (column 1) or not (column 2)
  - limiting to vacant cruises with the propensity score of turning on AI in [0.1, 0.9]

Model Fit



The figure compares the model prediction ("Survival") vs. actual data ("Kaplan-Meier curve")

Productivity gain by skills

	(1) Full	(2) Navi users	(3) Navi users $0.1 \leq PS \leq 0.9$
Navi usage $\times$ low-skilled	-0.076***	-0.059**	-0.081**
	(0.026)	(0.027)	(0.032)
Navi usage $\times$ high-skilled	-0.029	-0.038	-0.055
	(0.033)	(0.035)	(0.041)
$\log(p)$	0.181***	0.188***	0.248***
	(0.005)	(0.006)	(0.011)
driver FE	✓	✓	✓
ward FE	✓	✓	✓
date-hour FE	✓	✓	✓
N	63,426	28,909	6,567

- Observed only for the low-skilled (8%) with no gain on the high-skilled in column 1.
- The similar patterns are observed in
    - sample restriction (column 2), and trimming by the PS (column 3)

Productivity gain by skills: tertiles

	(1) Full	(2) Navi users	(3) Navi users $0.1 \leq PS \leq 0.9$
Navi usage $\times$ skill index 1st tertile	-0.077***	-0.064**	-0.081**
	(0.030)	(0.030)	(0.033)
Navi usage $\times$ skill index 2nd tertile	-0.071*	-0.058	-0.076
	(0.039)	(0.042)	(0.058)
Navi usage $\times$ skill index 3rd tertile	0.000	-0.019	-0.042
	(0.038)	(0.039)	(0.038)
$\log(p)$	0.181***	0.188***	0.248***
	(0.005)	(0.006)	(0.011)
driver FE	✓	✓	✓
ward FE	✓	✓	✓
date-hour FE	✓	✓	✓
N	63,426	28,909	6,567

- The similar pattern holds at tertiles.
  - No effect on the high-skilled.

Productivity gain by skills: Navi compliance

	(1) Full	(2) Full	(3) Navi users	(4) Navi users
Navi usage × low-skilled	-0.077***	-0.079**	-0.064**	-0.066**
	(0.030)	(0.031)	(0.030)	(0.032)
Navi usage × middle-skilled	-0.071*	-0.074*	-0.058	-0.060
	(0.039)	(0.041)	(0.042)	(0.043)
Navi usage × high-skilled	0.000	-0.005	-0.019	-0.025
	(0.038)	(0.039)	(0.039)	(0.041)
Navi usage × Navi compliance rate		-0.061		-0.062
		(0.052)		(0.052)
$\log(p)$	0.181***	0.181***	0.188***	0.189***
	(0.005)	(0.005)	(0.006)	(0.006)
driver FE	✓	✓	✓	✓
ward FE	✓	✓	✓	✓
date-hour FE	✓	✓	✓	✓
N	63,426	63,426	28,909	28,909

- The results are unchanged after we control for Navi compliance rate

How about fare per ride?

Dependent variable: $\ln(fare)$	(1) Navi users	(2) Navi users	(3) Navi users $0.1 \leq PS \leq 0.9$
Navi usage	-0.013		
	(0.013)		
Navi usage $\times$ low-skilled		-0.031	-0.031
		(0.019)	(0.026)
Navi usage $\times$ middle-skilled		-0.017	-0.026
		(0.018)	(0.019)
Navi usage $\times$ high-skilled		0.020	-0.011
		(0.024)	(0.029)
driver FE	✓	✓	✓
ward FE	✓	✓	✓
date-hour FE	✓	✓	✓
N	21,713	21,713	5,431

- Impact of AI on fares is not statistically significant.
- The magnitude of reduction in fare for the low-skilled is small (3%), *not offsetting* the reduction in search time (about 8%).

Instrumental variable approach (ongoing)

- We address the remaining concern of the endogeneous switching on/off of AI via two instrumental variables (ongoing).
- **IV1: Location of dropping off the previous customer**
  - The destination of the previous customer is arguably randomly determined.
  - Dropping off a customer at an *unfamiliar* location induces the driver to turn on AI.
- IV2: Past AI navi usage
  - Drivers accumulate the experience of AI navi usage by the past random events.
  - The past usage of AI navi induces the current AI navi usage.

Construction of the unfamiliarity index

- We construct *unfamiliarity* index to the location by (1-share).
- Suppose the drive i has 1,000 vacant cruises in *Oct/Nov* (pre-trial period), and the number of the starting vacant cruises (=dropping off the previous customer) at wards 1, 2, and 3 are 700, 200 and 100. (Note: In reality, we have 18 wards).
  - The share of wards 1, 2, and 3 are 0.7, 0.2 and 0.1.
  - For the vacant cruise starting at wards 1, 2 and 3 in *December* for drive i, we assign 0.3, 0.8 and 0.9.
- **Relevance:** It is likely that this driver turns on AI Navi when the starting vacant cruise at ward 3 (=0.9) than ward 1 (=0.3).
- **Exclusion restriction:** The destination of the previous customer will not affect the length of the following vacant cruise time other than through turning on AI (after controlling for starting ward FEs).

Instrumental Variable Approach (ongoing)

- We address the concern of the endogeneous switching on of AI via two instrumental variables.
- Location of drop off of the previous customer
  - The destination of the previous customer is randomly determined.
  - Dropping off a customer in an unfamiliar place induces the driver to turn on AI.
- **IV2: Past AI navi usage**
  - Drivers accumulate the experience of AI navi usage by the past quasi-random events.
  - The past usage of AI navi induces the current AI navi usage.

First Stage

- Probit regression of Navi Usage dummy on potential instruments.
- Both IVs seem strong.

Dependent variable: Navi usage dummy	(1) Navi users	(2) Navi users	(3) Navi users	(4) Navi users
1 - Past share of starting ward	0.421***		0.413***	0.553***
	(0.159)		(0.159)	(0.176)
# of navi usage		0.007***	0.007***	0.012***
		(0.002)	(0.002)	(0.003)
(1 - Past share of starting ward) $\times$ # of navi usage				-0.007**
				(0.003)
driver FE	✓	✓	✓	✓
ward FE	✓	✓	✓	✓
date-hour FE	✓	✓	✓	✓
N	26,021	26,021	26,021	26,021
N of drivers	201	201	201	201
χ^2-test	-	-	$\chi(2) = 18.05$	$\chi(3) = 28.33$
p-value	-	-	0.0001	0.00000

IV estimation via control function (ongoing)

We implement IV estimation using the control function approach.

First stage:

$$AI_{ijhs} = 1(Z_{ijhs-1}\gamma + v_{ijhs} > 0), \quad v_{ijhs}|Z_{ijhs-1} \sim N(0,1)$$

Generalized residual capturing the “endogeneous” part of AI_{ijhs}:

$$\begin{aligned} \hat{GR}_{ijhs} &= AI_{ijhs}\hat{E}(v_{ijhs}|AI_{ijhs}=1) + (1-AI_{ijhs})\hat{E}(v_{ijhs}|AI_{ijhs}=0) \\ &= AI_{ijhs}\frac{\phi(Z_{ijhs-1}\hat{\gamma})}{\Phi(Z_{ijhs-1}\hat{\gamma})} - (1-AI_{ijhs})\frac{\phi(Z_{ijhs-1}\hat{\gamma})}{1-\Phi(Z_{ijhs-1}\hat{\gamma})}, \end{aligned}$$

where Φ is CDF and ϕ is PDF of the standard normal distribution.

Second stage:

$$
\begin{aligned}
S_{ijhs}(t) &= \exp(-\lambda_{ijhs}(t) \cdot t^p) \\
\lambda_{ijh}(t) &= \exp\{-p(\alpha \cdot \text{AI Navi usage Dummy}_{ijhs,t} + \text{driver FE}_i \\
&\qquad + \text{ward FE}_j + \text{date-hour FE}_h + \mathbf{f}(\hat{\mathbf{GR}}_{\mathbf{ijhs}}))\}
\end{aligned}
$$

(i:driver, j:ward, h:date-hour, s:vacant cruise, t:time of vacant cruise)

Conclusion and Discussion (1/2)

- We study the impact of AI on productivity in the context of taxi drivers.
  - AI Navi improves productivity on average by 5%.
  - The impact concentrated on low-skilled drivers (8%)
- The results show that AI is *substitute* to skill *within* an occupation, echoing *across*-occupations study (e.g., Webb 2020).
- AI adoption may alter skill requirement for an occupation.
  - If demand-forecasting skill (or prediction skill in general) become less important, skills that cannot be easily automated (e.g., social skill; Deming and Kahn 2018) may become more important for taxi drivers.

Conclusion and Discussion (2/2)

- How valid is our finding beyond taxi drivers?
  - The result will be invalid even for taxi drivers once it is replaced by self-driving car with demand-forecasting AI.
  - But automating a task is costly under non-regularized environment, and the cost may be higher than wage saving (Autor, 2015).
  - The *partial* automation of tasks within an occupation (like our case of taxi drivers) is likely to persist in many occupations.
- Our finding can be applicable to other occupations in which the core skill involves a prediction task. For example, the likely candidates are
  - low-skilled paralegals benefited by legal tech AI that helps them review contracts to identify unusual clause,
  - low-skilled pathologists benefited by diagnostic imaging AI that detects malign tumors, than high-skilled counterparts.