Learning by Driving:
Evidence from taxi driver wages in Singapore[1]

Sumit Agarwal*, Shih-Fen Cheng+, Jussi Keppo* and Satyanarain Rengarajan*

October 2015

Abstract

The significance of "Learning by doing" (LBD) in the economy has been disputed by economists. We use a unique dataset from a Singaporean taxi fleet consisting of 3,250 drivers with over 520 million data points that track cabdrivers' minute-by-minute work routines for two years to test if taxi drivers exhibit LBD. After controlling for individual-level differences, i.e. socio-economics, we find strong evidence of LBD. We document this mechanism and show taxi drivers learn through temporal (when) and locational (where) and technology (booking) optimization. The study has implications for the contribution of LBD behaviour to the economy and its direct impact on individual wages. We also discuss the implication of our findings for taxi markets experiencing a rise in use of technology (i.e. booking systems based on applications), which has recently drawn the interest of policymakers.

Keywords: Learning-by-doing, Wages, Labor Supply, Taxi Drivers

JEL Classification Codes: J22, B49

[1] We benefited from the comments of Gene Amromin, Souphala Chomsisengphet, Nagpurnanand Prabhala, Tomek Piskorski, Wenlan Qian, Tarun Ramadorai, Amit Seru, and seminar participants at the National University of Singapore and Singapore Management University.

*National University of Singapore
+Singapore Management University

1. Introduction

Studies have shown that rises in productivity are higher with cumulative output particularly during early stages of production (Dutton, Thomas, & Butler, 1984; Jovanovic & Nyarko, 1995). This phenomenon is said to occur when producers learn from experience and cumulative output is a good proxy for that experienced (Thompson, 2001). Engineers refer to this as the "start-up curve"; economists call it the "learning curve" or "learning by doing" (LBD). Because of the difficulty in isolating the direct impact of LBD on productivity from changes in other inputs such as capital, labor, and technology, the economics literature is so far divided, with authors either supporting or opposing the contribution (and significance) of learning by doing to productivity growth.[2] In addition, identifying the channels of learning becomes more cumbersome as focus in the literature often seems to be more on the empirical validity of testing the learning-by-doing hypothesis.

There are several reasons why the prior debate is still ongoing. The first is the obvious existence of multi-collinearity between economies of scale and LBD as output increases: the learning curve and economies of scale go often together in any expanding firm and are statistically indistinguishable (Lamoreaux et al., 2007). A second persistent problem in finding evidence for LBD is the presence of confounding variables; it is not clear whether the identified productivity growth observed is due to LBD or increases in capital, change in labor, technology, organizational arrangements or decrease in product quality that is reflected as productivity growth in the data.

This could be solved by looking for historical events where investments (i.e., labor, capital, and technology) are stalled over long periods of time to test the impact of LBD on productivity. However, this situation is rare. Even in historical cases where such a lack of investment occurred (David, 1974; Lazonick & Brush, 1985; Lundberg, 1961), some authors argue changes in product quality, technology, or organization may have more of an influence on increasing productivity thereby diluting the contribution of LBD reported earlier. These difficulties in identification have often thwarted progress in identifying important channels through which individuals (not firms as a whole) infer policy across different industries.

Ideally, evidence supporting the LBD without any confounding factors would either require information on all aspects of the chosen firm/industry, including detailed, long-term records of capital inflow/outflow, labor inflow/outflow, technology and product quality changes as control variables. The lack of such datasets is the third problem hindering a compelling argument for the significance of LBD on productivity and ultimately wages of individuals. Consequently, the combination of confounding variables and data requirements prevents

[2] Initial studies supporting the Learning By Doing (LBD henceforth) hypothesis showed favourable evidence in the manufacturing sector that involves the production of both raw materials and machinery (Alchian, 1963; Paul A. David, 1973; Genberg, 1992; Lazonick & Brush, 1985; Lieberman, 1984; Lundberg, 1961; Wright, 1936). More recently there has been a slew of studies debunking earlier findings lowering or even nullifying the contribution of LBD to productivity growth (Bell & Scott-Kemmis, 1990; Lamoreaux, Raff, & Temin, 2007; Mishina, 1999; Sinclair, Klepper, & Cohen, 2000; Thompson, 2001) while others find weak support (Levitt, List, & Syverson, 2012) .Some recent papers have found evidence for LBD in finance as well (Kandel, Ofer, & Sarig, 1993; Pissarides, 1997; Seru, Shumway, & Stoffman, 2010).

useful investigation on how and at what rate individuals learn, and at which point, if any, learning is bounded.

Our paper sheds light on LBD literature not only by showing strong empirical support for the presence of LBD in the taxi industry and its direct impact on daily wages of workers, but also identifying channels of learning exhibited by individuals through their driving behaviour. We differ from earlier studies that look for evidence in manufacturing industries and we turn to a service-oriented transportation industry: the taxicab marketplace in Singapore.

We follow 3,250 taxi drivers in Singapore each hour of the day and every day of the year for nearly two years (23 months), giving us detailed information about their activity such as trips taken, origination, destination, working/idle time and wages earned. This unique dataset has a wealth of information on individuals' socio-economic variables (e.g. race, age, and education levels). Our dataset also has detailed driving behaviour with accurate time and location stamps, with over 520 million observations, that give us the ability to compare drivers and determine their channels of learning.

Our primary motivation for this study comes from observing a simple trend of the progression of taxi driver wages (aggregate) in Singapore during the period 2009-2010. Figure 1 below plots net wages of single-driven[3] taxi drivers (3,250) in Singapore over time for twenty three months. This might not necessarily mean the entire population is "learning" to perform better in the marketplace; indeed, the standard deviation in Figure 1 (in red) shows that there is a wide spectrum of wage growth or decline under which taxi drivers may fall.

[Insert Figure 1 here]

The steady increase in aggregate average wages during this period might suggest other causes. We started with a simple question: On average do drivers earn more with more days spent driving? To test this, we first run a simple regression in Table 1 with wages on total number of Days Driven (DD) by each taxi driver[4]. As expected, we observe the coefficient (DD) to be positive and significant. For each extra day taxi driver spends on the road he stands to gain S$0.028 on his daily net fare. While this might seem small, incremental gains at the individual and aggregate level are significant over time. For example, the average driver working for an extra one year would increase his daily average wage from the S$82.46 to S$92.68 (+13.39 %) and within two years this increases to S$102.90 (+24.79 %), which translates into additional annual income of S$3,679 and S$7,358 respectively. One possibility for this increase could be due to growing external demand over time for taxi rides. However, the relationship is stable even after removing seasonality using daily, monthly, yearly-fixed effects, albeit with a smaller coefficient as seen in models (2)-(4), ruling out the possibility that increasing wages are solely due to increasing demand.

[**Insert Table 1 here**]

[3] Taxis in Singapore are rented out by companies to individuals which can be either single or multiple-driven (shared). We study only single driven taxis, shared behavior would make our analysis intractable.
[4] Our sample drivers are not working every day; we also have new entrants/exits during this period that gives us balanced answer to this question.

Another possible cause of the increase in wages could be an increase in labor participation: drivers might on average spend more time looking for passengers over time. However, we see an opposite pattern in our descriptive statistics. Daily labor force participation (hours spent working) declined by 31 hours (-4 %) from 2009 to 2010 (see Table 3). That taxi drivers are, in general, learning to perform better with experience or "learning by driving," is the intuitive conclusion once other possibilities are eliminated.

Why do we believe taxi drivers in Singapore are exhibiting learning behaviour? There are two simple reasons. First, wages-per-kilometre earned remained constant throughout our sample period[5]. This is important as it eliminates wage inflation, one of the biggest confounders of detecting learning behaviour using wage data. Secondly, Singapore remains one of the few countries with a very low barrier to participate in the taxi market.[6] Taxi drivers in Singapore are given minimal training, which spans for only few weeks (as opposed to approximately 2 years in London). Given the short training period, most Singaporean taxi drivers are highly incentivised by a dynamic and competitive environment to learn "on the job" in order to maximize their daily wages.

In addition our study period of two years is a short enough time for any major capital investment for such an industry;[7] we suspect this is an advantage as it minimises any long term advantages gained by drivers through explicit capital expenditure, such as new cars or technology improvements that may reflect in higher wages. Studies that focus on manufacturing activity find these investments are difficult if not impossible to weed out. In addition, the increasing use of machinery in the manufacturing sector inhibits opportunity of employees exhibiting LBD. On the other hand, LBD still plays a more prominent role in many service sectors, making it an ideal environment to study this phenomenon.

We claim that taxi drivers who make better decisions consistently (i.e. those who are learning) will experience increased wages over time after controlling for changes in external demand (using time-fixed effects) and internal differences (socio-economics). During the period of study, there are no changes in prices, technology, or competition[8], yet wages increased even after controlling for changes in demand over time. By process of elimination, we attribute wage gains over time to improvement in driving behaviour.

In Figure 1, given the high standard deviation (in red), theoretically we can expect the slope to be higher for some drivers, and stagnant or declining for others. To understand how drivers learn and perform better in the marketplace, we divide our sample into groups which we identify as "progressive" (learners), "regressive" and "stagnant" groups (non-learners) using a simple autoregressive model (AR) model (Section 4.1) and study their driving behaviour (section 4.2). Having categorised the drivers, we proceed to the main part of the paper. What

[5] See "Yearbook of Statistics Singapore 2015", page 250, under "Transport/Taxi Fares", available at http://www.singstat.gov.sg/docs/default-source/default-document-library/publications/publications_and_papers/reference/yearbook_2015/yos2015.pdf
[6] The low barrier refers to becoming a taxi driver, not starting a company with its own fleet.
[7] Every driver used the same car throughout the study period (identified by vehicle ID).
[8] The number of cabs on the road fluctuates seasonally; however, there is no linear pattern that would explain the increase or decrease in drivers' wages over time.

are the general determinants of wages of taxi drivers in Singapore? How important are these determinants for progressive drivers as compared to other regressive and stagnant drivers? Are there specific channels (variables) through which they learn?

We identify three important determinants that have an impact on daily wages of taxi drivers in all groups. The first and most important is choice of time of day. In Singapore, peak hours comprise the heavy commuter traffic of the working population into the monocentric city, making them a profitable time slot. The second important factor is the ability to obtain passenger trips via several channels; the booking channel (bookings through a taxi company's call center) provides the greatest potential wage growth, followed by increasing labor supply (daily hours worked) and the cruising channel (street pickups). Other factors such as socio-economic differences (race, marital status and education levels) play only a modest role in explaining differences in daily wages among drivers. Minor differences in daily wages due race persist where being Indian or Chinese rather than Malay (base) seems to contribute to a wage premium.

Taxi drivers with consistent wage growth (the progressive group) generally exhibit higher returns on trips obtained via the booking and cruising relative to other groups. In general, we find that an one-unit increase in passenger trips obtained, normalized by labor supply, via the booking or cruising channel increases progressive drivers' daily net wages by 16% and 12.6% respectively over our two-year study period. In contrast the taxi drivers with declining wages (the regressive group as we identify them), have higher returns on purely mechanical channels such as labor supply and strategic labor supply (labor supply during high demand periods of the day), which are relatively easier ways of finding passengers. For example, the wage premium on driving taxis during peak period is higher for drivers in the regressive and stagnant groups as compared to the progressives who have gained an upper hand overall by learning to obtain wage premiums on trips obtained through the booking service or cruising channels.

Our contribution to the academic literature is two-fold. First, we show strong support for LBD in taxi drivers and its direct impact on individual wages. While previous studies have measured its effects on an aggregate level (i.e., plant or firm's total output per hour), here we show direct evidence of learning or skills on daily wages, marking the importance of LBD in the taxi industry. We bypass the problem of confounders affecting LBD by studying a service industry.[9] Previous studies focussing heavily on manufacturing industry to demonstrate the existence and benefits of LBD on increasing productivity have recently come under criticism due to presence of confounding variables and low data quality. In contrast using a straightforward identification, we demonstrate a case for LBD in a service industry: the competitive Singaporean taxicab marketplace.

Secondly and most importantly, we identify specific channels of learning behaviour to support the previous claim, minimizing the impact of other confounders in our identification of LBD among taxi drivers. Specifically, we find evidence of drivers increasing their wages through four channels: labor supply, cruising trips, booking trips, and strategic labor supply.

[9] To our knowledge we are the first to select a service industry to study LBD activity among workers.

We show how progressive drivers differ in usage of these channels relative to stagnant and regressive taxi drivers in Singapore. Previous studies have typically focused on disputing the existence and/or significance of learning by doing at the manufacturing plant or firm level. We also extend the research by showing mechanisms through which LBD is achieved among taxi drivers after controlling for socio-economic differences.

The remaining part of this paper is organized as follows: Section 2 reviews the literature on Learning by Doing (LBD). Section 3 gives a brief description of our data, followed by a simple theoretical model of LBD in the taxi industry. Section 4 details our identification strategy and empirical modelling techniques. Section 5 discusses the results, and Section 6 concludes the study with future avenues of research.

2. Learning by Doing

We can define LBD as a phenomenon where the average cost of unit production decreases as cumulative output over time increases under a fixed labor-capital-land ratio and a fixed output. This cost reduction is achieved by "labor learning"[10] and not by other factors such as economies of scale, product standardisation, and new technology.

Although LBD is intuitive, literature surrounding the importance or contribution of learning by doing on productivity and output is highly contradictory. As mentioned earlier, this is because the effect is difficult to isolate; in other words, quantifying residual that is uncorrelated with other variables without additional costs (e.g., labor and capital) is still a challenge. Also, LBD is truly costless if working time is fixed and not substituted for leisure (Killingsworth, 1982). To see LBD's true contribution, its impact needs to be controlled for working time.

Wright's (1936) study of the U.S. airframe industry found evidence of a learning curve that effectively increased industry performance during World War II. He noticed that as the quantity of units (or product) manufactured doubles, the direct number of hours used to produce a unit decreases at a uniform rate. These claims were later supported by similar studies of airframe production in the U.S. by Asher (1956) and Alchian (1963).

This influential idea was followed by a stream of studies showing evidence for LBD in different manufacturing industries. For example, Montgomery (1943) showed a sharp learning curve in a shipbuilding exercise wherein labor requirements had reduced three-fourth over five years. In economics literature, this phenomenon became known as the "Horndal effect" through Erik Lundberg (1961)'s work on productivity and profitability of a public steel plant in Horndal, Sweden where he observed a consistent increase in productivity in spite of the absence in investments.

Arrow (1962), reflecting on his theoretical work on the economic implications of LBD, later supported Lundberg's findings, saying that the productivity growth at Horndal could "only be imputed to learning from experience."

[10] This type of learning could be as simple as choosing to be at the right place and at the right time. It could also be driver's decision-making processes such as choosing short or long trips during peak/non-peak hours.

Over the last several decades, empirical studies have supported the notion of LBD contributing to productivity increases when labor and capital are constant in other industries. David (1973) found evidence for the Horndal Effect in the U.S.: a Lawrence, Massachusetts, cotton textile mill increased labor productivity about 2% per annum (1836-1856) despite the absence of new investments in machinery. Using data on 37 chemical plants (and products) in the U.S, Lieberman (1984) deconstructs the various reasons[11] for increased output and finds strong evidence for learning. As in previous studies (Rapping, 1965; Sheshinski, 1967), Lieberman (1984) finds cumulative output a better predictor of learning than time alone; using time as a proxy for "experience" might therefore result in a less accurate picture of LBD in regressions.

Recently, there have been critical studies on the findings reported in the earlier empirical work which lend only partial or no support to the LBD hypothesis. In a critical review of David (1973)'s study on the cotton mill, Lazonick and Brush (1985) argue that the productivity increase may not be solely LBD, but may be primarily due to pre-existing social-economic relationship advantages between employees and employers in the plant. Bell and Scott-Kemmis (1990) provide qualitative evidence against LBD as a predominant factor behind productivity growth in airframe and shipbuilding industries during wartime in the U.S. Instead they attribute it to numerous other factors (e.g. "unfixed" facilities, pre-existing human capital, improved materials, and economies of scale resulting from machinery).

Mishina (1999) revisits the historical case study of airframe manufacturing at a Boeing plant where the Horndal Effect was observed by Alchian (1963). He dismisses LBD as an important factor that led to significant decreases in labor hours during output expansion. Instead he cites increases in capital investments and ensuing systemic organizational adjustments as main drivers of productivity growth. Sinclair et al. (2000) conducts an in-depth analysis of 221 specialty chemical plants using qualitative and quantitative data and finds that productivity growth can be largely attributed to variations in research and development rather than LBD.

While many studies dispute the existence and significance of LBD, to our knowledge there are no studies with datasets sufficiently rich enough to identify specific channels of learning. Past findings that support LBD are therefore subject to criticism on missing confounders. By contrast, our study uses detailed, individual-level, real-time taxi driver behaviour, going beyond supporting LBD to explore specific mechanisms through which taxi drivers learn in the marketplace and how this learning improves over time. This is our main contribution to the literature.

More specifically, in this paper we first ask whether taxi drivers experience LBD in Singapore. Secondly, we extend our analysis to show how they learn, and the direct impact of this learning on daily wages of taxi drivers in Singapore.

3. Data Sources, Empirical Design and a Simple Model of Learning by Doing

[11] Increasing output can signify learning as well as other economic forces in action such as "labour learning, process improvement, product standardization and economies of scale" (Lieberman, 1984).

3.1. Taxi Data on Daily Wages

The taxi industry in Singapore is highly regulated and operates in a unique way: only citizens more than 30 years old are allowed to drive taxis, and drivers can either rent taxis from one of seven approved taxi operators[12] or own their taxi (private taxi ownership is less than 10% of the total fleet). The rental cost is set by individual operator and usually depends on the brand and type of taxi. Rent covers all vehicle-related expenses such as insurance, maintenance, road tax, and vehicle tax; the only cost not covered is fuel, which is borne by individual drivers. Drivers keep all revenue earned and trip fees are constant during our sample period. Due to this revenue structure, revenue earned by individual drivers can be used as a proxy for driver skill.

We obtained a dataset containing the hourly wage data of individual taxi drivers[13] from a taxi company in Singapore for the period of 2009–2010. This hourly data is aggregated at daily and monthly levels. As fares from this period do not change, gross wage data can be directly computed from the raw data without adjustment. Net wage can then be inferred by subtracting operational costs, which include 1) variable fuel charge calculated by total distance travelled and 2) daily taxi rental costs. To correctly account for the rental cost for individual drivers, we include only taxis that are driven by single drivers.[14] When we mention taxis or drivers, we refer specifically to taxis driven by a single driver.

The distribution of daily net wages among drivers follows a uniform distribution (see Figure 2 below). In the raw data, a typical driver earns less than S$20 per day about 22.16 % of the days during the two-year period; in other words, daily income of taxi drivers in Singapore is not chronologically consistent, showing the presence of seasonality in passenger traffic.

[Insert Figure 2]

Our dataset includes several important variables associated with wages. The first is related to labor supply and indicates duration spent on the roads (hours) and the total number of passenger trips undertaken each day. Labor supply can be further disaggregated into daytime and nighttime, which reflects drivers' choices on when to provide their services.[15] Similarly, the number of passenger trips (trip count) can be disaggregated into what we call "trip count (cruising)" and "trip count (booking)." Trip count (cruising) refers to trips that result from

[12] For detail rules on becoming a taxi driver, see: http://www.lta.gov.sg/content/ltaweb/en/public-transport/taxis/industry-matters-for-taxi-drivers/driving-a-taxi-in-singapore.html. Private taxis are otherwise called yellow top taxis.

[13] Identified by unique, anonymized driver ID.

[14] Drivers in Singapore who rent taxis are allowed to recruit one or more secondary drivers to share the driving time as well as rental cost; this information is unfortunately not available in the dataset. Therefore, to accurately estimate the cost component, we choose to only look at single-driver taxis.

[15] For each day in the dataset, a driving time is split into two equal periods: 6am to 6pm is referred to as "daytime" while 6pm to 6am is referred to as "nighttime."

street pickups, while trip count (booking) refers to trips that originate from pre-arranged bookings through a taxi company's call center[16].

The dataset also contains drivers' self-reported socioeconomic variables such as age, race, driving experience (number of years since attaining vocational driving license), gender, marital status (married, unmarried, divorced) and education level. In summary, the raw dataset contains 10,345 anonymized drivers of single-driven taxis labeled with a unique driver ID; there are 2,815,654 raw observations for span of 23 months (January 2009 till December 2010, except December 2009, which was missing). A description of all the socio-economic variables in our data is given in Table 2.

[Insert Table 2]

In general, we see that the more trips a driver takes the higher his daily wages. Figure 3 above plots the average net income earned by taxi drivers for a given trip count which ranges from 0-80. Typically, a taxi driver breaks even if he makes approximately 10 trips in a given day; for the net income of S$20, he would need to take slightly more than 12 trips per day.

[Insert Figure 3]

The raw data was cleaned to remove outliers and erroneous records. First, for each driver we remove days without any driving hours (including holidays); without this removal the sample would be skewed towards a zero net fare. Then we removed drivers who did less than one month of driving during the two-year period as they are not representative for our study. Next, there were instances in the data where trip count was zero but total fare was a positive number; we believe these are computer-generated errors, and we removed them. To further remove potential outliers, we removed records that are among top one percentile and bottom one percentile in terms of 'productivity' (wages/total number of hours worked) in the sample. Finally, the sample was further restricted to drivers belonging to three major race groups: Chinese, Indian and Malay; in aggregate more than 91% of all records belong to these three major race groups. After this restriction, the racial composition among single-driven taxi drivers is 72 % Chinese, 11% Malay, and 8% Indian, roughly consistent with the overall racial composition of Singaporean citizens.

We construct additional variables. For example, we record trips taken during high demand periods of the day (strategic labor supply) to be used as a control variable. We computed the total number of trips taken during these periods using hourly dataset by identifying peak hour information[17] and matched with the daily database using driver ID for each month, day and

[16] In Singapore, a potential rider can book a taxi for immediate pickup or several hours in advance by contacting the fleet's call center via voice call or text message. Upon receiving the booking call, the call center would send this booking request via mobile data terminal to a small number of selected vacant taxis (usually the few closest taxis within certain radius). The driver who responds quickest via the mobile data terminal gets the job. For pre-booked trips, riders pay a fixed booking fee, which starts at S$2.5 and is more expensive during rush hours.

[17] Peak hours were computed by observing high demand (above mean) hours in a day for the entire two-year sample.

year. The final sample consists of 3,816 drivers with 1,083,914 observations; we give the descriptive statistics of the cleaned sample divided into three social strata in Table 3. The drivers are divided into a low-income group (bottom 15^{th} percentile or two-year average net daily wage S$20), a high-income group (top 15^{th} percentile or two-year average net daily wage S$121), and a middle-income group (S$20-121).

[Insert Table 3]

There are some key differences observed when we split Table 3 by race. Table 4 gives the split by Chinese, Indian and Malay taxi drivers in Singapore. We observe that in these two years, Indians, on average, have higher net wages (S$91) as compared to Chinese (S$81) and Malay drivers (S$64); however, Indians also seem to drive more (12.81 hours) as compared to Chinese and Malay drivers (12.6 and 12.56 hours respectively). On average, Chinese taxi drivers seem to have more experience as measured by years on their driving license (30.63) compared to Indian (25.88) and Malay (25.13) populations. The data suggests an increase in overall productivity over time, i.e., all races in general earned higher average wages and worked less in 2010 as compared to 2009.

[Insert Table 4]

3.2. Determinants of Taxi Drivers' Wages

To identify determinants for high wages, we first run a regression model for the log of wages by controlling for socioeconomic heterogeneity among drivers. To control for daily variations in demand we introduce a day-level fixed effect variable (τ_t). The regression specification is represented as follows,

$$Log\left(net\ wages_{i,t}\right) = \alpha \quad + \beta_0 \cdot labor\ supply_i + \beta_1 \cdot strategic\ labor\ supply_i + \beta_2 \cdot strategic\ hours_i + \beta_3 \cdot socio_i + \tau_t + \varepsilon_{i,t}\ , \tag{1}$$

where t and i are used to denote time and driver indices respectively, α is an intercept term, and $\varepsilon_{i,t}$ is an *i.i.d.* error term. β_0 is the coefficient of the *labor supply* variable, which measures the mechanical relationship between time spent on the road for each driver i at each time interval t (day). β_1 is the coefficient of *strategic labor supply* that is the trips taken during high demand periods of the day. The importance of other socioeconomic variables such as age, experience, education, and race is also observed in this exercise.

3.3. Labor Skills and Wage Disparity

The ability to learn can be due to two factors. The first is an improvement in memory: drivers tend to accumulate knowledge on hotspots in the city such as major tourist spots, and office and residential spaces with frequent riders. The second channel is more dynamic and requires knowledge about best time of day and best days of the week. One may learn this through social interactions with other drivers in the city, reading newspapers or listening to radio. Both of these factors are important for learning.

Woollett and Maguire (2011) is the most well-known related study from the medical literature that provides evidence that taxi drivers indeed learn from their driving experience. After confirming the neurological evidence that links spatial navigation and memorization skills (which is critical in passing the exam for taxi license in London) to the volume of gray matter in the hippocampus, the authors further showed in a longitudinal study that taxi trainees who manage to pass the license exam saw a significant growth in the gray matter volume in the hippocampus, the region of the brain known to be responsible for memory and spatial navigation. While memorizing spatial information certainly helps in navigation, it is merely one of many factors that might contribute to the productivity of taxi drivers. In fact, as argued by Camerer, Babcock, Loewenstein, and Thaler (1997) and Varakantham, Cheng, Gordon, and Ahmed (2012), deciding when and where to drive seems to have much stronger impact on a driver's performance.

One direct consequence of skill acquisition for taxi drivers is faster wage growth and higher wage premiums over time (Glaeser & Mare, 1994). Thus, it is safe to assume that drivers who consistently experience wage growth are also acquiring new skills (i.e. learning) after controlling for external (environmental) factors or internal differences (individual characteristics). In the context of Singapore, we equate consistent wage growth with learning behavior. Thus, in order to observe learning behavior using our data, we simply ask which drivers experience wage growth, and whether there are drivers whose wages stagnate or decline over time? Before we attempt to answer these questions using our identifications strategy (Section 4), the next sub-section provides a simple economic model for learning for taxi drivers.

3.4. Simple Model of Learning by Doing

Let us consider the case of a single taxi driver. The driver maximizes the expected daily profits. That is,

$$E[\textstyle\sum_{i=1}^{N} F_i] - C = E[N]E[F_i] - C, \qquad (2)$$

where N is a random number of trips per day, F_i is the random fare received for the i^{th} trip, N and F_i are independent, and $C = r + f$ is a daily cost that includes both fuel (f) and rental costs (r). We use $E[N]$ and $E[F_i]$ to denote the expected values of random variables N and F_i.

We assume that r is a constant[18], to reflect the fact that taxi rental cost stays unchanged during the course of study. Fuel costs are variable, but can be approximated using total hours driven by each driver on each day. For a driver to improve his expected daily profit, he can

[18] This constant empirically depends on the brand of the vehicle a taxi driver rents.

try to increase $E[N]$ and/or $E[F_i]$; and operationally speaking, this can be achieved by the following three practical channels: increasing working hours (this raises $E[N]$), finding better working hours (this raises $E[N]$ and/or $E[F_i]$), and improving his customer-finding skills to find more customers or customers who would take longer trips (this raises mainly $E[N]$, but could potentially raise $E[F_i]$), as well). These are the most important channels of learning that we analyze using our empirical models.

While the first channel is mechanical, the second and the third channels require drivers to learn from their experience. In the next section, we use a range of statistical approaches to identify driver's learning (the latter two channels above) for increasing profits.

4. Identification Strategy

4.1. Autoregressive (AR) Modeling

As a first step, we identify drivers whose wages grow, stagnate, or decrease over time. Once drivers are labeled, we then identify common characteristics for those who learn and those who do not. To do this, we employ an autoregressive model of order one (AR(1)) for each driver by regressing current daily net income ($net\ wages_t$) on the wages in the previous day ($net\ wages_{t-1}$) for the entire study horizon from 2009 to 2010.

$$net\ wages_{i,t} = \alpha \quad + \beta_i * \ net\ wages_{i,t-1} + \varepsilon_{it}, \tag{3}$$

where t and i are used to denote time and driver indices respectively. In Eq. (3) above, β_i is the respective AR(1) coefficient for driver i in the labor force, and ε_{it} is the error term of the driver at time t (day), where $i \in \{1,2, \dots 3816\}$. The resultant set of coefficients (βs) for all taxi drivers are split into three major groups based on the sign of the coefficient and significance levels (10%).

We label drivers who exhibit significant wage increases over time, as illustrated by their positive and significant βs, as the *progressive group* (there are 1,874 of them). We label drivers who exhibit significant wage decreases over time, as illustrated by their negative and significant βs, as the *regressive group* (there are 107 of them). We group all other drivers with positive (or negative) yet insignificant βs, into the *stagnant group*, since they exhibit no significant trends in wage changes during our sample period (there are 1,835 of them). Figures 4, 5 and 6 below plot the distributions of βs from the AR(1) model of three respective groups.

[Insert Figures 4, 5 and 6]

The separation of these groups based on the AR(1) model helps us investigate the systemic differences in their behavior over time. The descriptive statistics for each group in Table 5 show some stark differences between taxi drivers who have wage growth, decline and stagnation during our sample periods.[19]

[Insert Table 5 (a), 5 (b) and 5 (c)]

4.2. OLS Regressions

We use a simple measure to capture skill acquisition among taxi drivers that we call *labor efficiency B/C*, which is computed by dividing the total number of trips taken in each day (*trip count*) via booking or cruising by the total number of hours worked (*labor supply*). We also use *strategic labor supply* which refers to passenger trips obtained during peak hours of the day divided by the total number of hours worked (*labor supply*). It is one of the ways in which drivers can earn more as fares charged for their service are higher during peak hour periods. Next, we estimate the following regression:

$$net\ wages_{i,t} = \alpha \quad + \beta_0 \cdot labor\ supply_{i,t} + \beta_1 \cdot labor\ efficiencyB_{i,t} + \beta_2 \cdot labor\ efficiencyC_{i,t} + \beta_3 \cdot strategic\ labor\ supply_{i,t} + \beta_4 \cdot socio_i + \tau_t + \varepsilon_{i,t}, \tag{4}$$

where α is an intercept term and ε_{it} is an *i.i.d.* error term. β_0 is the estimated coefficient for *labor supply*. β_1, β_2, and β_3 are the estimated coefficients of *labor efficiency B/C* and *strategic labor supply*. The coefficients $\beta_1 - \beta_3$ are our key parameters in analyzing the learning channels. For instance, if $\beta_1 > \beta_2$ and $\beta_1 > 0$, then the implication is that drivers in that particular group are exhibiting high productivity via the booking channel, and its magnitude reflects the level of importance in explaining wages. Finally, β_4 captures wage differentials due to socio-economic differences between taxi drivers; we include some driver characteristics such as age, age squared, race, experience, education level and marital status for this purpose. τ_t is the daily time-fixed effect, which we use to remove the influence of special days (holidays or unusually busy days). We run the specification above separately on the three groups of taxi drivers identified using the AR(1) model, namely progressive, regressive and stagnant groups.

Although we use a number of control variables to eliminate differences in worker characteristics, there is a possibility of omitted variable bias in the specification given in Eq. (4). To strengthen our specification, we include a driver-level fixed effect (λ_i) in Eq. (5). The regression model is thus modified:

$$\log\left(net\ wages_{i,t}\right) = \alpha \quad + \beta_0 \cdot labor\ supply_{i,t} + \beta_1 \cdot labor\ efficiencyB_{i,t} + \beta_2 \cdot labor\ efficiencyC_{i,t} + \beta_3 \cdot strategic\ labor\ supply_{i,t} + \lambda_i + \tau_t + \varepsilon_{i,t} \tag{5}$$

[19] We estimate the correlation between the beta and income and find the correlation to be -0.039 with a p-value 0.002. This implies that low income drivers have a higher beta; in other words, there is mean reversion. However, in our regression analysis, we control for wages.

The coefficients $\beta_1 - \beta_3$, for each of the three groups help us compare the relative importance of each learning channel among the three groups by its impact on daily wages of Singaporean taxi drivers.

4.3. Time Series Analysis

So far we have pointed out the existence of LBD and channels of learning for taxi drivers in Singapore. In this section, we aggregate the data and show monthly behavior of key variables among the three groups that shows a macro-picture of LBD between groups. Specifically, we examine the monthly response of trip count and net wages of taxi drivers over the two-year period. We further investigate the dynamic evolution of taxi drivers' behavior within the progressive group in terms of trip count and its impact on monthly net wages by conducting in-sample heterogeneity tests. The model specification is straightforward that looks at the monthly effect on wages and trip count over the 23 months (the max that our data allows). Unlike the previous analysis that assumes that the coefficient is constant over time, here we allow for a flexible functional form – testing for non-linearity in the wage and trip counts over time. These results are shown in Figures 8-10.

5. Results

5.1. OLS Regressions

Table 6 summarizes the results from the basic regression using Eq. (4), estimating the various determinants of wages for taxi drivers in Singapore. The single most important determinant of wages seems to be the ability of drivers to choose more profitable peak hours (strategic labor supply). After controlling for socio-economics and seasonal variations (daily, monthly and yearly), each extra trip (per peak hour) increases their wages by S$53 per day.

The second most important determinant seems to be driver's reliance on the booking or cruising channels. Daily net wages of taxi drivers go up by S$37 for those who rely on booking as compared to S$28 for trips obtained through cruising on the streets after controlling for seasonal and individual level differences (Model 2). Labor supply is the third most important channel that has a significant impact on net wages: the longer one drives the greater the increase in trips and wages. As the base Model (2) shows, initially for every extra hour spent on the road, the driver receives a S$30 increase in his daily wages after controlling for daily, monthly, and yearly-fixed effects. However, this relationship follows a U-curve as the variable *labor squared* is negative and significant, suggesting declining returns for extra time spent on the roads.

Apart from these major factors, race, age and education also have an impact on wages. We find that taxi drivers belonging to the Indian and Chinese community in Singapore earn daily

S$11 and S$4.3 respectively more than Malay community. Older drivers in general do slightly better than younger drivers; however, this relationship is non-linear and significant (*agesq*). Each year to taxi driver's age increases his/her daily net wages by S$0.26. Similarly and possibly a correlated variable marital status has a negative impact on net wages.

The relationship between education and wages earned by drivers is mixed. Drivers with a degree from a polytechnic or PRU make about S$4 more daily than drivers with only primary education. While having a secondary education helps drivers as compared to only a primary school education, for e.g. drivers with ITE[20] education make 7.4 less than drivers with only primary education. Similarly those with university degrees do not necessarily make more money than others. In fact they make S$2 and S$7 less than drivers with only primary education.

[Insert Table 6]

We want investigated whether the progressive group differs from the others groups in the return on skills learned. Tables 7(a) and 7(b) report the importance of labor supply, labor efficiency and strategic labor supply on daily wages for all three groups (progressive, regressive, and stagnant) as identified by the AR(1) model in subsection 5.1. The following regressions shown in Tables 7(a) and 7(b) are estimated using Eq. (4) and Eq. (5) respectively.

[Insert Table 7 (a)]

After controlling for time fixed effects in Model (2) in Table 7(a), we find significant differences among the groups on the returns related labor supply, labor efficiency (booking/cruising), and strategic labor supply, suggesting why some perform better than others. For example, the progressive group had the highest return on booking as well as cruising trips as compared to stagnant and regressive groups. After normalizing for labor supply, each extra booking trip obtained per hour (labor efficiency) increases the drivers daily wage by 16% for progressive drivers as compared to 10.8% and 9.5% for stagnant and regressive groups. Similarly, each extra trip obtained per hour via cruising on the streets increases daily wages of progressives by 12.6% as opposed to only 8.9% and 5.5% for stagnant and regressive drivers respectively. There are two more mechanical ways regressive drivers learn. Firstly, their returns on daily wages relative labor supply (13.2 %) is slightly higher than the progressive (12.6 %) and stagnant groups (12.4 %) of drivers, meaning they earn more by staying longer on the streets. Secondly, their increased returns on daily wages attributed to obtaining trips during peak hours are much higher than drivers belonging to the progressive and stagnant groups.

For each additional trip obtained by the regressive group during peak hours normalized by labor supply, daily wages increase by 26% compared to 20.8% and 20.9 % for drivers in

[20] Institute of Technical Education system set up by the Ministry of Education (MOE) in Singapore that provides pre-employment training to secondary school leavers and continuing education and training to working adults.

progressive and stagnant groups. The progressives seem to learn by emphasizing booking and cruising trips whereas the regressive group depends on two mechanical channels to earn more wages - labor supply and strategic labor supply, i.e., peak hours. The stagnant group is similar to the progressives in terms of their learning behavior but their skill level or returns from each channel is lower if we look at the two-year window. These patterns are very consistent albeit with lesser magnitude even after introducing stringent controls for each individual driver using fixed-effect models as shown in Table 7(b).

[Insert Table 7 (b)]

For robustness[21], we repeat the regressions carried out in Tables 6 and 7 using two sub-sample periods, 2009 and 2010, to test the structural validity of the previously estimated coefficients and behaviour over time. Tables 8(a) and 8(b) repeat earlier regression Models (1) and (2) from Table 7(a) and 7(b) estimated using Eq. (4) and Eq. (5).

[Insert Table 8 (a)]

We observe that returns on labor supply increase for all three groups with similar magnitudes from 2009 to 2010. For the labor efficiency via booking and cruising variables, the magnitudes drop from 2009 to 2010 for all the three groups; however, the differences outlined earlier between the groups remain the same. There is a possibility that the observed differences in magnitudes of our key variables are partially seasonally driven. After controlling for individual differences (driver-fixed effects), the regressive group registers the lowest returns for labor efficiency via the cruising channel, suggesting low skill at choosing local passenger pickup spots relative to other groups. On the other hand, progressives learn via the booking and cruising and strategic labor supply channel to make consistent profits.

[Insert Table 8 (b)]

5.2. Time Series

We report the cumulative changes of trip counts and net fare for different groups in Figure 7 (*trip count* in Panel A and *net fares* in Panel B). In the first few months (1-4) we find very modest gains in trip counts compared to the base month (0, i.e. January 2009), which improves gradually and consistently for the next eighteen months. Regression estimates for stagnant drivers show a somewhat cyclical pattern with lowering trip counts during the first eleven months and some gains made at the beginning of second year only to dip again. The regressive group follows very similar patterns; however, it fails to recover in the second year. Most point estimates of stagnant and regressive groups are, however, statistically insignificant in contrast to the progressive group which is statistically significant at the 1%, 5% or 10% levels. Thus, as the confidence interval in Figure 7 (Panel A) suggests, only the

[21] In our sample, we observed that some taxi drivers worked only for one particular year (2009 or 2010). This might be a problem if many drivers are skewed towards the end of the sample period (2010) which may get reflected in aggregate analysis showing learning behavior when in fact it is due to the distribution of the drivers over time. For robustness, we removed these single-year represented individuals and re-ran our regressions. The results remain by and large unchanged. Those tables can be provided on request.

cumulative estimates for the progressive group are statistically significant, demonstrating the learning ability of drivers.

[Insert Figure 7]

Regression estimates for progressive, stagnant and regressive groups show an upward curve in net wages (Figure 7, Panel B) over the 22-month period; however, the gains made by progressive group are much higher compared to the stagnant and regressive groups (i.e. $Progressive_{net\ wages} > Stagnant_{net\ wages} > Regressive_{net\ wages}$), consistent with the outcome of the trip count behavior in the previous regressions.

5.3. Heterogeneity in Time Series

Figures 8-10 plot the cumulative changes of trip counts (Panel A) and net fare (Panel B) by race, marital status and education level. Even though we see strong evidence supporting the learning behavior among the progressives, this pattern is not homogenous across the socio-economic divisions of taxi drivers. While previous regressions show a static picture of the differences in learning along socio-economic lines, in this section we extend them to provide a more dynamic context. As expected, we find considerable heterogeneity in learning patterns as evidenced their trip count and corresponding net fares obtained over time along racial lines, marital status and education qualifications, after controlling for temporal and individual-fixed effects.

For example, we see major differences between races: the Chinese population of progressive drivers, when compared to otherwise similar Indian and Malay populations, shows strong support for learning behavior as evidenced by a steady increase in trip count and corresponding increases in income (*net fare*) throughout the 22-month period (see Figure 8, Panels A and B).

[Insert Figure 8 here]

Among the progressives of the Indian population, we notice something interesting: the monthly trip counts do not increase consistently over time (Figure 8 Panel A); however, their net wages steadily improve over the same period (Figure 8 Panel B). This indicates that they are either picking niche hours to drive (e.g., night-time trips charge extra in surcharges) or choosing locations that often yield longer trips (e.g., airports or tourist spots). Either way they are learning, but this learning seems to differ from the Chinese drivers who seem to achieve similar wage growth by picking up more passengers. The small sample size of minority populations does not allow us to explore this phenomenon in detail.

When progressive drivers are divided along marital lines we do not observe any stark differences among these groups, but we do notice married drivers make more trips over time, which is statistically significant at 1% level. While the progression of net wages has a continuous upward inflection for the married, single and divorced groups (see Figure 9, Panel A and B), it is consistent and significant only for the single and married drivers.

[Insert Figure 9 here]

Finally, an important variable that seems to be significant in our static regressions is education level of drivers. Therefore, in Figure 10 we further analyze the learning behavior in progressives holding a primary, secondary or university education.[22] Our cumulative estimates from dynamic regressions show that progressive drivers with primary or secondary[23] degree consistently increase their monthly trip count and net wages over our study period. There is considerable deviation in trip count and net wage behavior among progressives with a university degree rendering it difficult to attribute any learning behavior by these individuals.

[Insert Figure 10 here]

6. Conclusion

Studies have shown that the rise in productivity is higher with cumulative output particularly during early stages of production. This phenomenon is said to occur when producers or workers learn from their experience, and cumulative output is a good proxy for that experience. Economists refer to this phenomenon as learning by doing (LBD). Empiricists have been divided: early findings support the impact of LBD on output whereas more recent studies either debunk of show weak evidence of this contribution. Most of this controversy is due to the difficulty in isolating the direct impact of LBD on wages among other confounders such as changes in other inputs such as capital, labor, and technology over time.

This paper exploits microdata on labor supply and daily wages from taxi drivers to show evidence for the LBD hypothesis. In our paper, we bypass some of the challenge of isolating the LBD effect from the influence of confounders by exploiting a sub-section of the Singaporean transportation sector: the highly competitive taxi marketplace. We first show strong support for learning by doing behavior among a sub-section of taxi drivers in Singapore. After that we show how taxi drivers learn, in other words the channels (i.e. labor supply, labor efficiency and strategic labor supply) through which taxi drivers increase their daily wages.

We use two methods to demonstrate credible learning experience among a sub-section of taxi drivers. First we use a simple AR model on daily wages to show a section of drivers

[22] These refer to driver's highest qualification. Drivers with these three types of education level represent over 97% of the sample. We discarded drivers with other education backgrounds (e.g. Poly, ITE etc.) due to low sample size for dynamic regressions.

[23] The error bars however are considerably small as compared to the drivers with only a primary school degree.

exhibiting consistent wage growth relative to others. We then run dynamic regressions on three sub-sections based on their wage growth to plot differences in performance over time.

In general, we find that drivers learn and have the ability to improve their earning potential through three important channels, namely labor supply, labor efficiency, and strategic hour selection in decreasing order of importance. We also find that other factors such as socio-economic differences (race, marital status and education levels) only play a very modest role in explaining differences in wages between these groups.

Progressive taxi drivers, or those with consistent wage growth, in general exhibit higher returns on trips obtained specifically via the booking and cruising channels relative to other groups. We find that a one-unit increase in passenger trips obtained normalized by labor supply, via the booking or cruising channel increases daily net wages for progressive drivers by 16% and 12.6% respectively over our 2-year study period. By contrast, the taxi drivers with declining wages, the regressive group, have higher returns on purely mechanical channels such as labor supply and strategic labor supply which are relatively easier ways of finding passengers.

The labor efficiency channel of picking up passengers (booking & cruising) is less effective for drivers in this regressive group. A one-unit increase in passenger trips obtained via the booking and cruising channel normalized by labor supply increased daily net wages by only 9.5% and 5.5% respectively (almost half of the progressive group's magnitude). For those taxi drivers whose wages were either stagnant or highly cyclical (what we refer to as stagnant group), we find that labor efficiency via booking or cruising is very modest compared to the progressives showing room for more improvement. These estimates remain consistent even after robustness tests.

Our main contribution is twofold. First, we show strong support for the LBD and its direct impact on individual wages. Prior literature in this area generally focuses on manufacturing plants to show or refute support for LBD; we diverge from these studies and look into a service sector that provides us a cleaner approach for identification purposes. Second, studies that find support for LBD often find it at an aggregate level such as total output per hour; here we not only show direct evidence of "learning" but also how individuals learn to improve their wages. This shows the role and important channels through which LBD plays in the taxi industry.

An ancillary finding was that, in addition to other channels, the progressive drivers (high skilled) rely disproportionately on trips generated through the automated booking service. We believe that the rise in mobile-based booking applications may increase wage inequality due to differences in technological know-how.

Finally, an alternate hypothesis can be made pointing out to the fact that some drivers earn consistently more than others (i.e. what we call learning in this paper) may not be due to learning but random events. One can overcome that possibility by pointing out to a series of studies in the medical literature which has found that navigational learning in taxi drivers changes the hippocampus region of the brain which is believed to play an important role in memory and spatial navigation. From those studies we know that such on-the-road learning can have a significant impact on driver's wages even over the short term.

An implication for policy makers dealing with wage inequality in the taxi market would be to conduct voluntary training programs in navigating routes, identifying passenger hot spots along important time slots of the day, and empowering and incentivising drivers to adopt new technology. This could be an important step in levelling the playing field.[24]

[24] To our surprise, we recently learnt that the Singaporean government has recently set up a 2.5 million dollar fund which will enable taxi drivers to "…upgrade their skills in customer service, safe driving and information technology" (see http://www.straitstimes.com/singapore/transport/taxi-drivers-to-get-training-support-with-252-million-fund). In line with the findings of this paper, we believe such steps by governments or companies could alleviate wage inequality among low skilled and high skilled individuals in the taxi marketplace.

References

Agarwal, S., Diao, M., Pan, J., & Sing, T. F. (2014). Labor Supply Decisions of Singaporean Cab Drivers. Working paper, available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.364.708&rep=rep1&type=pdf

Agarwal, S., Liu, C., & Souleles, N. S. (2007). The reaction of consumer spending and debt to tax rebates--evidence from consumer credit data. *Journal of Political Economy*, 115(6), 986-1019.

Agarwal, S., & Qian, W. (2014). Consumption and debt response to unanticipated income shocks: Evidence from a natural experiment in Singapore. *American Economic Review*, 104(12), 4205-4230.

Alchian, A. (1963). Reliability of progress curves in airframe production. *Econometrica: Journal of the Econometric Society*, 31(4), 679-693.

Arrow, K. J. (1962). The economic implications Of Learning By Doing. *The Review of Economic Studies*, 29(3), 155-173.

Asher, H. (1956). *Cost-quantity relationships in the airframe industry*. Research report, available at https://www.rand.org/content/dam/rand/pubs/reports/2007/R291.pdf

Bell, M., & Scott-Kemmis, D. (2010). The mythology of learning-by-doing in World War II airframe and ship production.*International Journal of Technological Learning, Research and Development*, 3(1), 1-35.

Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. (1997). Labor supply of New York City cabdrivers: One day at a time. *The Quarterly Journal of Economics*, 112(2), 407-441.

David, P. A. (1973). The “Horndal effect” in Lowell, 1834–1856: A short-run learning curve for integrated cotton textile mills. *Explorations in Economic History*, 10(2), 131-150.

David, P. A. (1974). The “Horndal effect” in Lowell, 1834–1856: A short-run learning curve for integrated cotton textile mills. *Explorations in Economic History*, 10(2), 131-150.

Dutton, J. M., Thomas, A., & Butler, J. E. (1984). The history of progress functions as a managerial technology. *Business History Review*, 58(2), 204-233.

Genberg, M. (1992). The Horndal effect: productivity growth without capital investments at Horndalsverken during 1927 and 1952. *Department of Economic History, University of Uppsala, Sweden*. Available at https://www.econbiz.de/Record/the-horndal-effect-productivity-growth-without-capital-investment-at-horndalsverken-between-1927-and-1952-genberg-mats/10000896236

Glaeser, E. L., & Mare, D. C. (2001. Cities and skills. *Journal of Labor Economics,* 19(2), 316-42.

Jovanovic, B., & Nyarko, Y. (1995). A Bayesian learning model fitted to a variety of empirical learning curves. *Brookings Papers on Economic Activity. Microeconomics*, 1, 247-305.

Kandel, S., Ofer, A. R., & Sarig, O. (1993). Learning from trading. *Review of Financial Studies,* 6(3), 507-526.

Kempf, E., Manconi, A., & Spalt, O. G. (2013). Learning by doing: The value of experience and the origins of skill for mutual fund managers. Working paper, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2124896

Killingsworth, M. R. (1982). "Learning by doing" and "investment in training": asynthesis of two "rival" models of the life cycle. *The Review of Economic Studies, 49*(2), 263-271.

Lamoreaux, N. R., Raff, D. M., & Temin, P. (2007). *Learning by Doing in Markets, Firms, and Countries*. Chicago: University of Chicago Press.

Lazonick, W., & Brush, T. (1985). The "Horndal Effect" in early US manufacturing. *Explorations in Economic History,* 22(1), 53-96.

Levitt, S. D., List, J. A., & Syverson, C. (2013). Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant. *Journal of Political Economy,* 121(4), 643-681.

Lieberman, M. B. (1984). The learning curve and pricing in the chemical processing industries. *The RAND Journal of Economics,* 15(2), 213-228.

Lundberg, E. (1961). *Produktivitet och räntabilitet: Studier i kapitalets betydelse inom svenskt näringsliv.* Stockholm: SNS.

Mishina, K. (1999). Learning by new experiences: Revisiting the flying fortress learning curve. In N. R. Lamoreaux, D. M. Raff, & P. Temin (Eds.), *Learning by doing in markets, firms, and countries* (pp. 145-184). Chicago: University of Chicago Press.

Pissarides, C. A. (1997). Learning by trading and the returns to human capital in developing countries. *The World Bank Economic Review,* 11(1), 17-32.

Rapping, L. (1965). Learning and World War II production functions. *The Review of Economics and Statistics,* 47(1), 81-86.

Seru, A., Shumway, T., & Stoffman, N. (2010). Learning by trading. *Review of Financial studies,* 23(2), 705-739.

Sheshinski, E. (1967). Optimal Accumulation with Learning by Doing. *Essays on the Theory of Optimal Economic Growth.* Cambridge, MA: M. I. T. Press, 31-52.

Sinclair, G., Klepper, S., & Cohen, W. (2000). What's Experience Got to *Do With It*? Sources of cost reduction in a large specialty chemicals producer. *Management Science*, 46(1), 28-45.

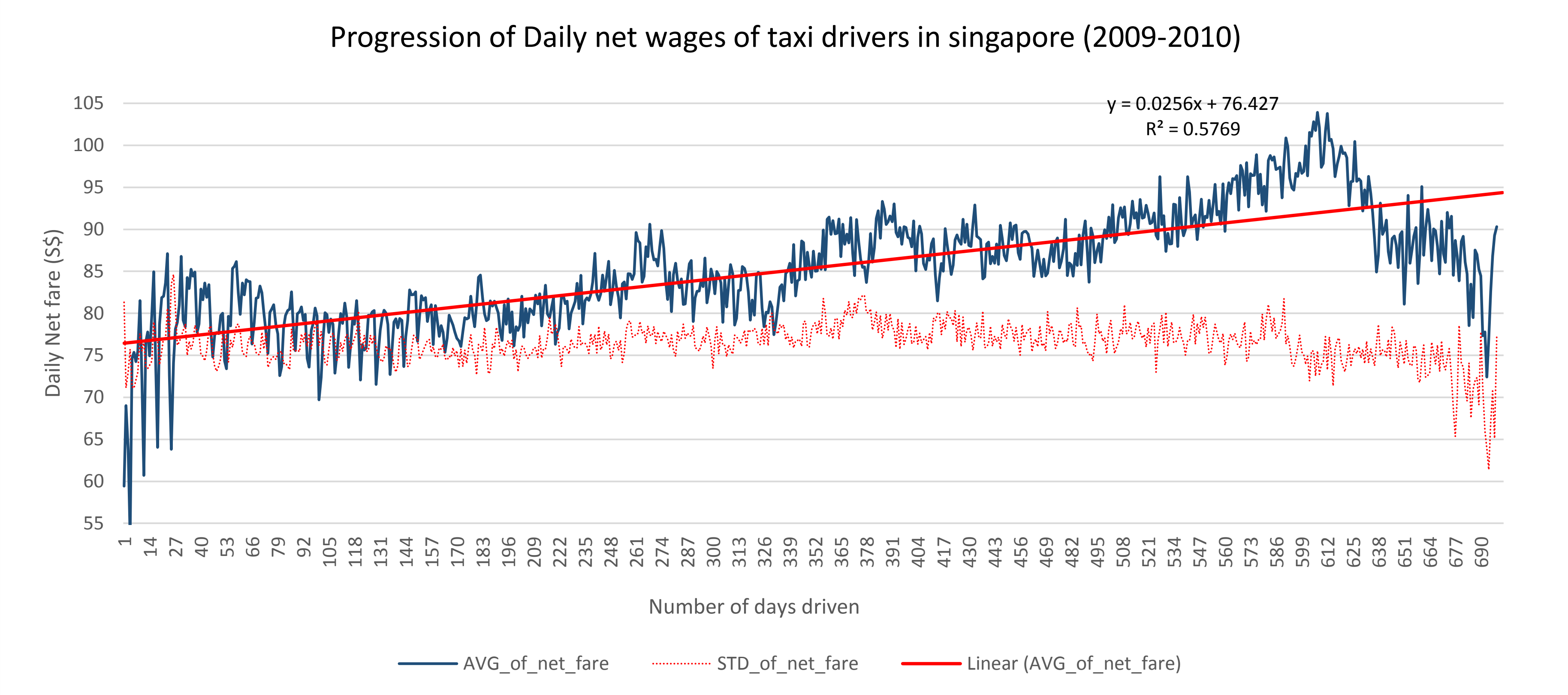
Thompson, P. (2001). How much did the liberty shipbuilders learn? New evidence for an old case study. *Journal of Political Economy,* 109(1), 103-137.

Varakantham, P. R., Cheng, S.-F., Gordon, G., & Ahmed, A. (2012). Decision support for agent populations in uncertain and congested environments. Working paper, available at http://ink.library.smu.edu.sg/sis_research/1567/

Woollett, K., & Maguire, E. A. (2011). Acquiring “the knowledge” of London's layout drives structural brain changes. *Current Biology,* 21(24), 2109-2114.

Wright, T. P. (1936). Factors affecting the cost of airplanes. *Journal of the Aeronautical Sciences (Institute of the Aeronautical Sciences), 3*(4), 122-128.

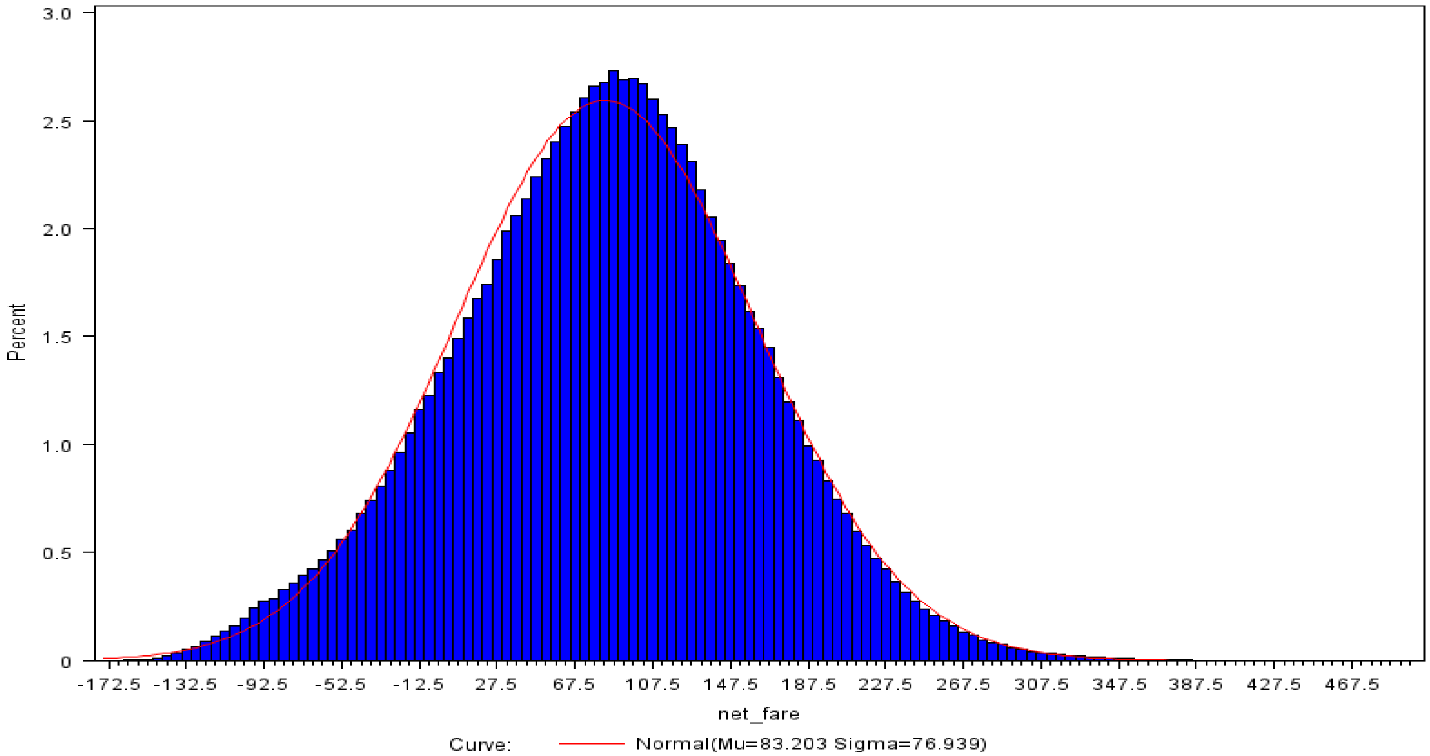
Figure 1: Aggregated movement of daily taxi fares (net) of drivers given by number of days driven.



Note: The horizontal column represents experience, i.e. number of days driven by each driver arranged in a linear fashion.

Figure 2: Distribution of daily taxi fares (net)

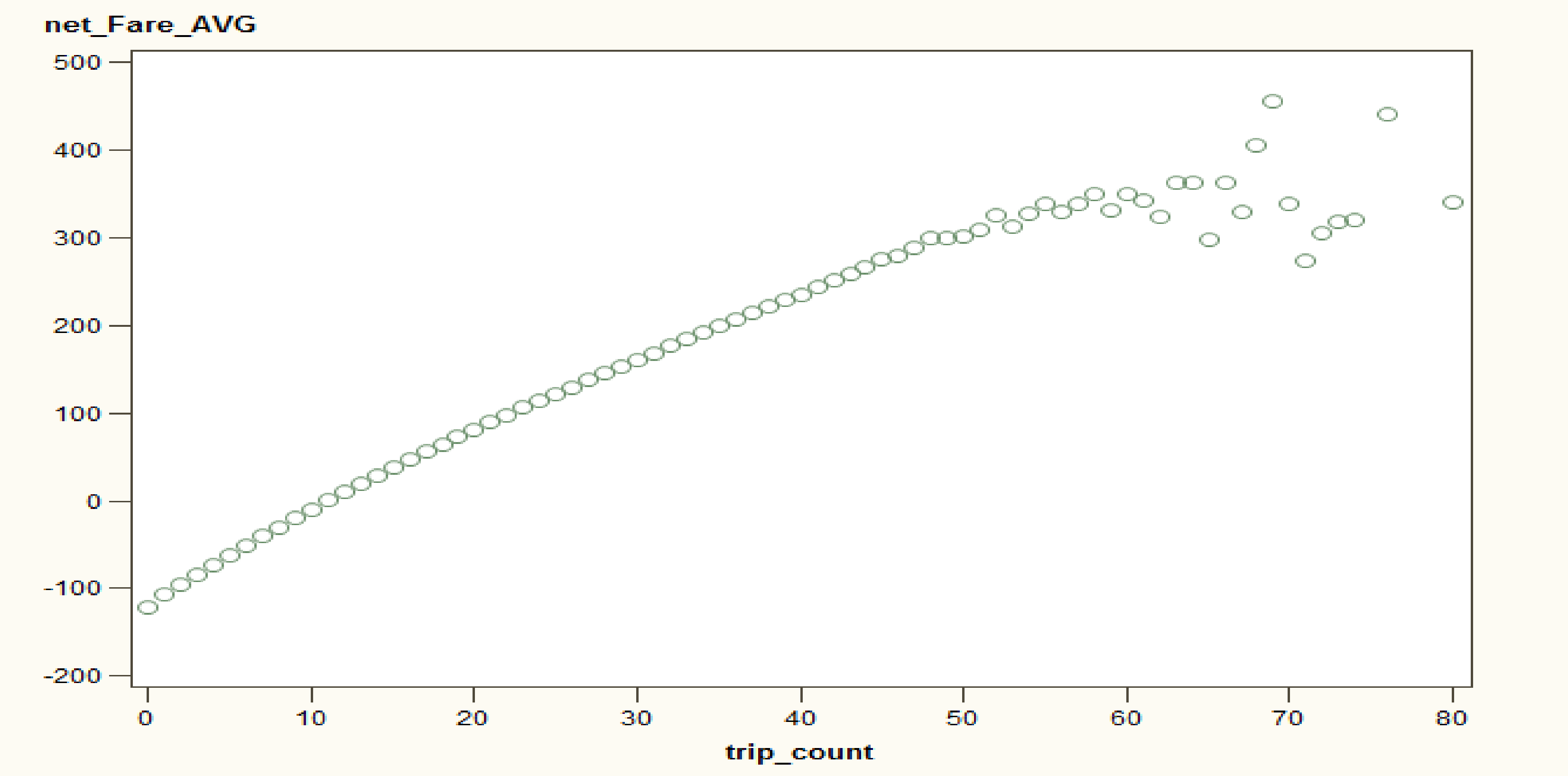
(No of obs. =1,083,929*)*



Note: The above figure plots the distribution of net fare earned by drivers every day for the period of 23 months from 2009 to 2010 in Singapore. The negative end of the distribution indicates days with low or no income when offset by daily-fixed costs such as rental and fuel.

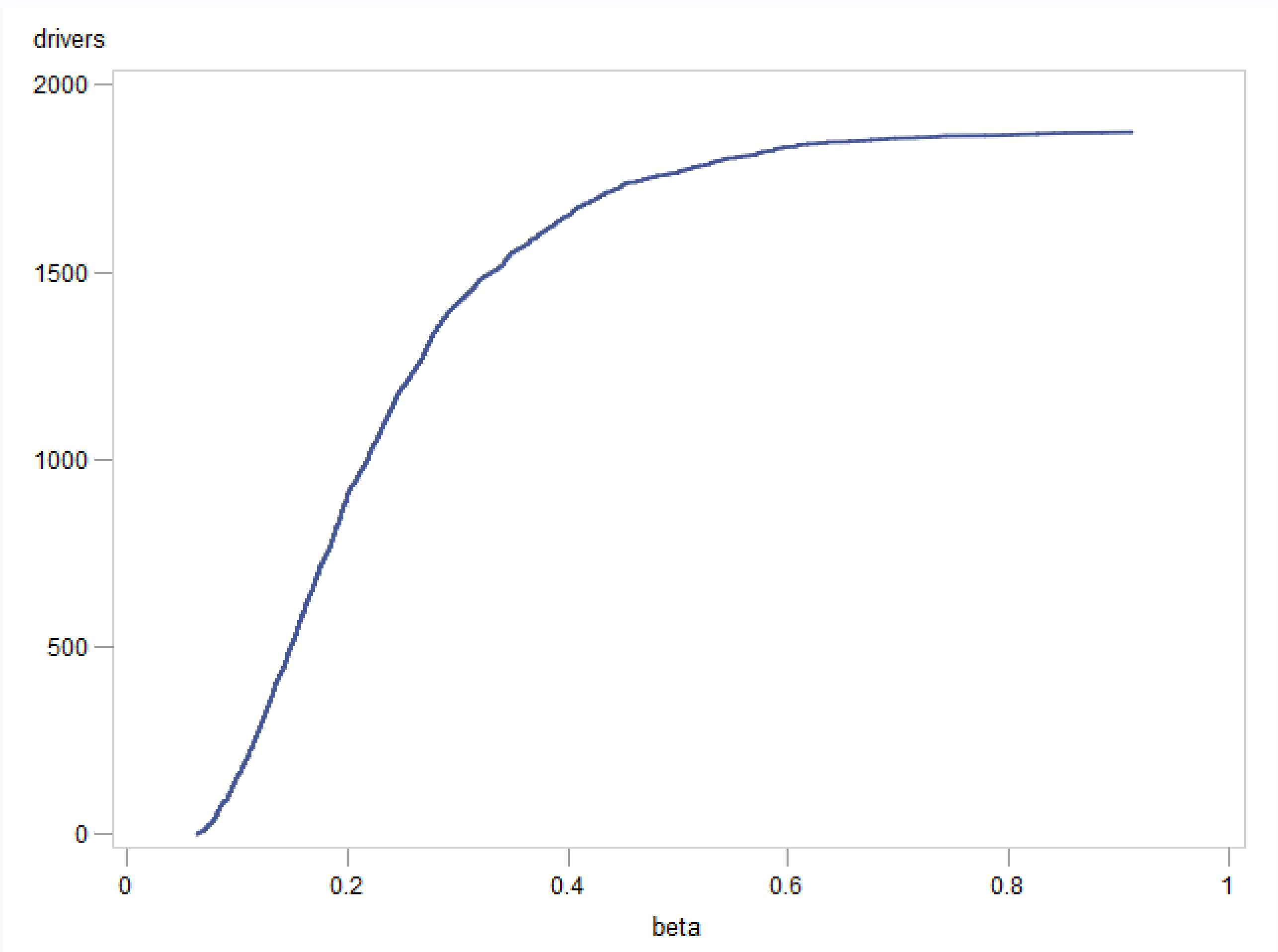
Figure 3: Average taxi fare (net) earned per day given by trip count

(No of obs. =1,083,929)



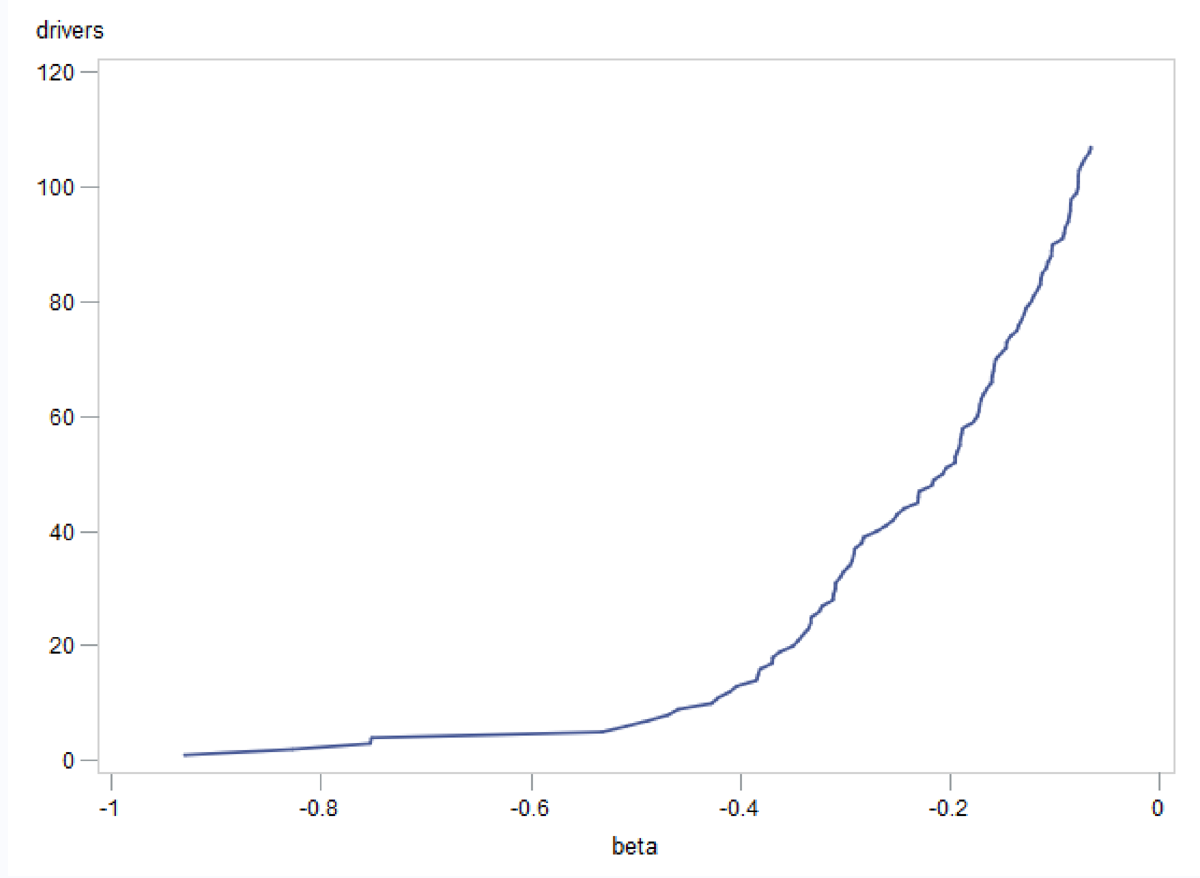
Note: Figure 3 plots the average net income (y-axis) earned by taxi drivers for a given trip count (x-axis) using the raw data.

Figure 4: Positive and significant Beta(s) given using AR(1) model (progressive group)



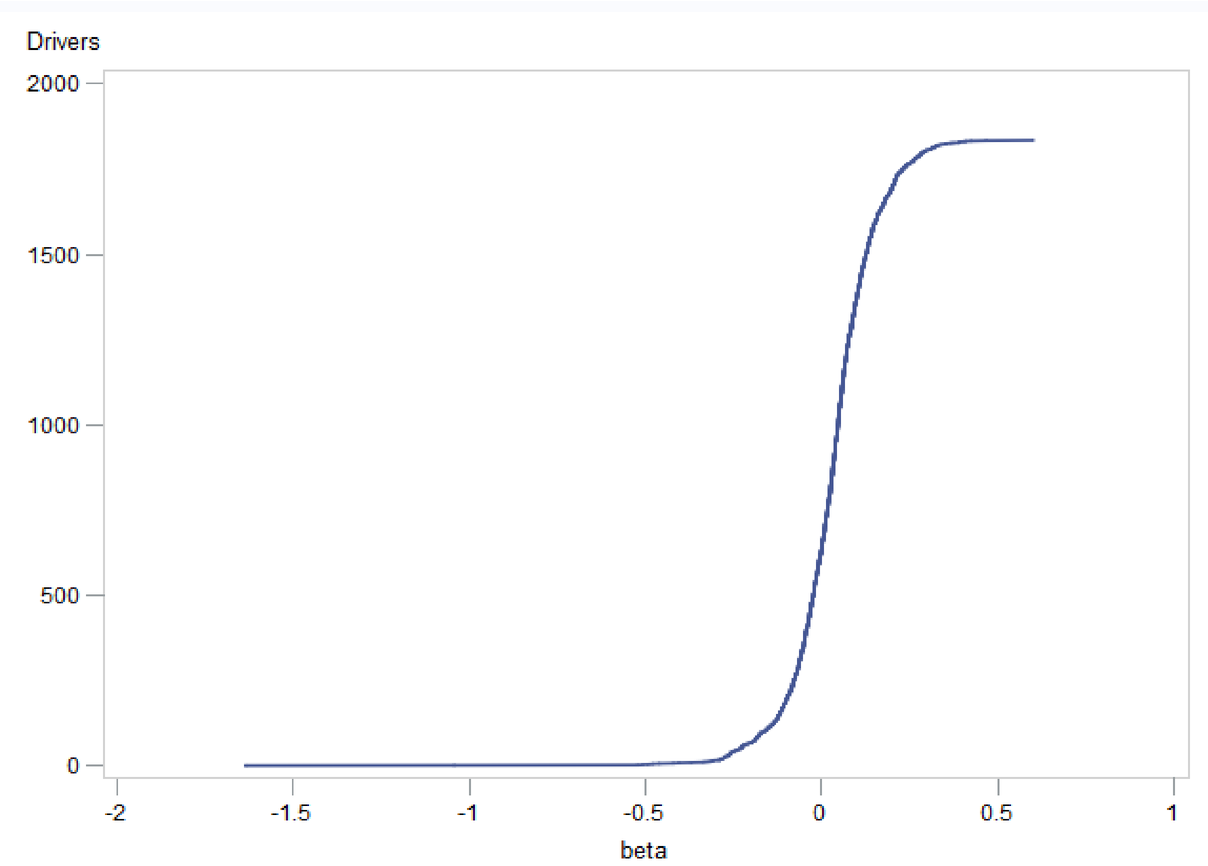
Note: Figure 4 plots the positive and significant coefficients from estimating the AR(1) model. The AR(1) model using specification given in Eq. (3) is estimated for all drivers in the sample (excluding low-income drivers).

Figure 5: negative and significant Beta(s) given using AR(1) model (regressive group)



Note: Figure 5 plots the negative and significant coefficients from estimating the AR(1) model. The AR(1) model using specification given in Eq. (3) is estimated for all drivers in the sample (excluding low-income drivers).

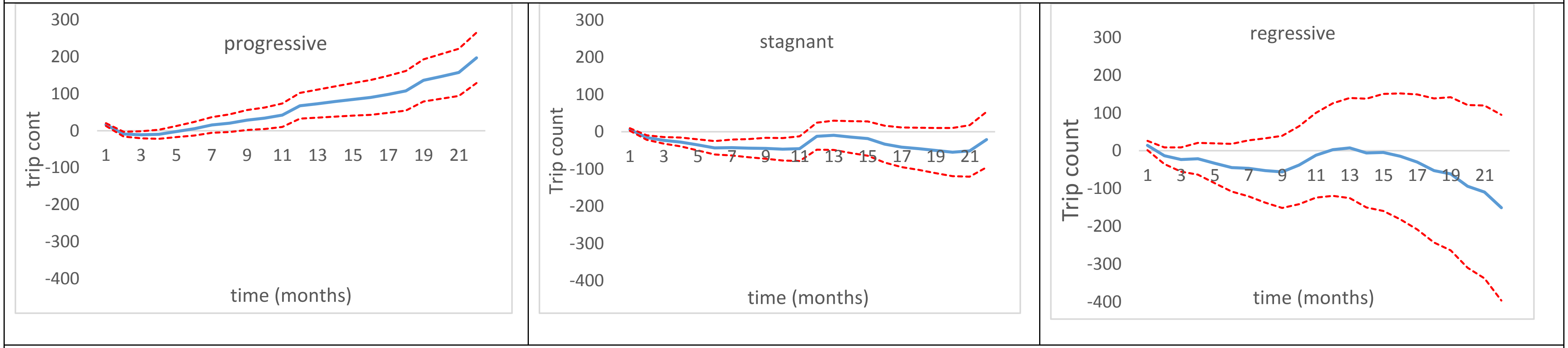
Figure 6: Positive/negative and insignificant Beta(s) given using AR(1) model



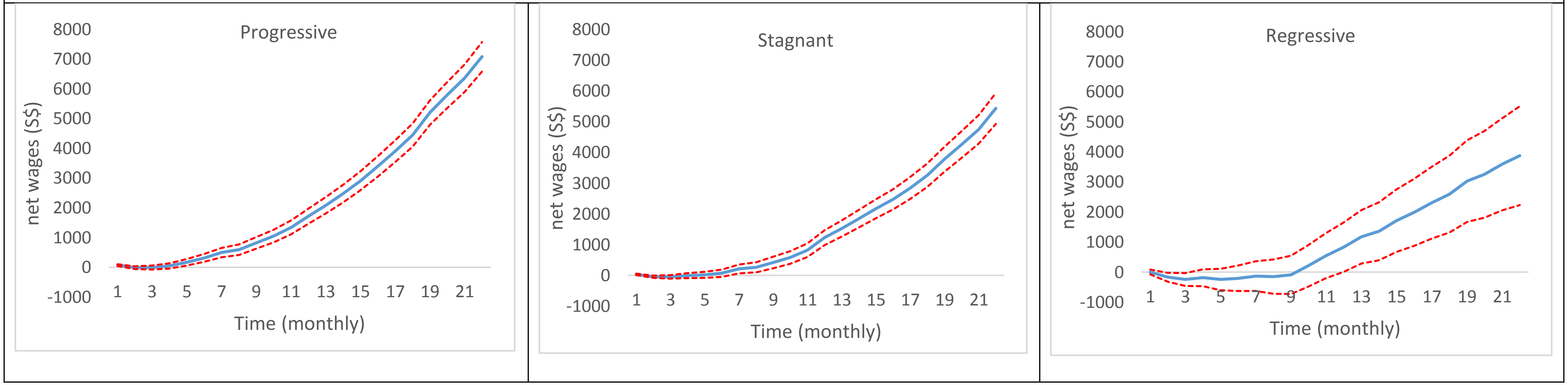
Note: Figure 6 plots the positive/negative and insignificant coefficients from estimating the AR(1) model. The AR(1) model using specification given in Eq. (3) is estimated for all drivers in the sample (excluding low-income drivers).

Figure 7: Cumulative change of taxi trip counts & net fare given by groups

Panel A. Dependent variable : Trip count



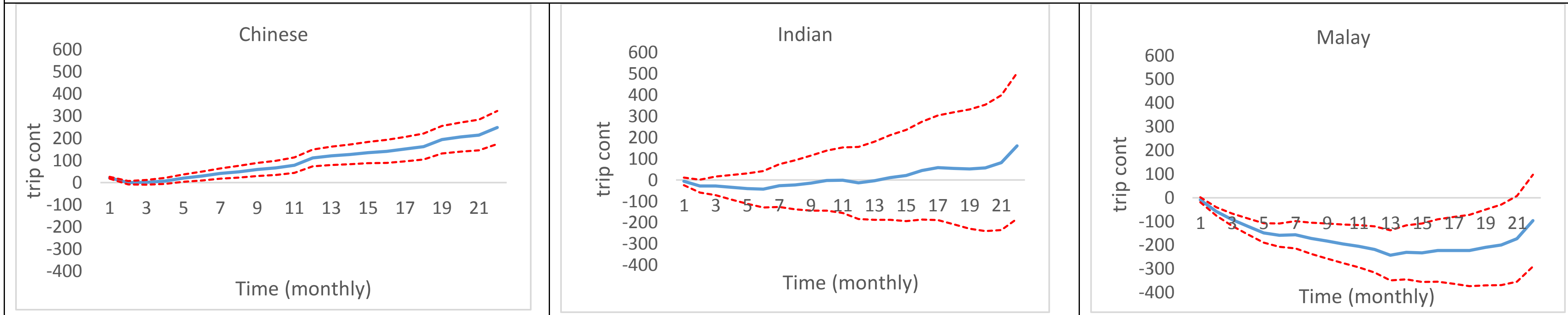
Panel B. Dependent variable : Net fare



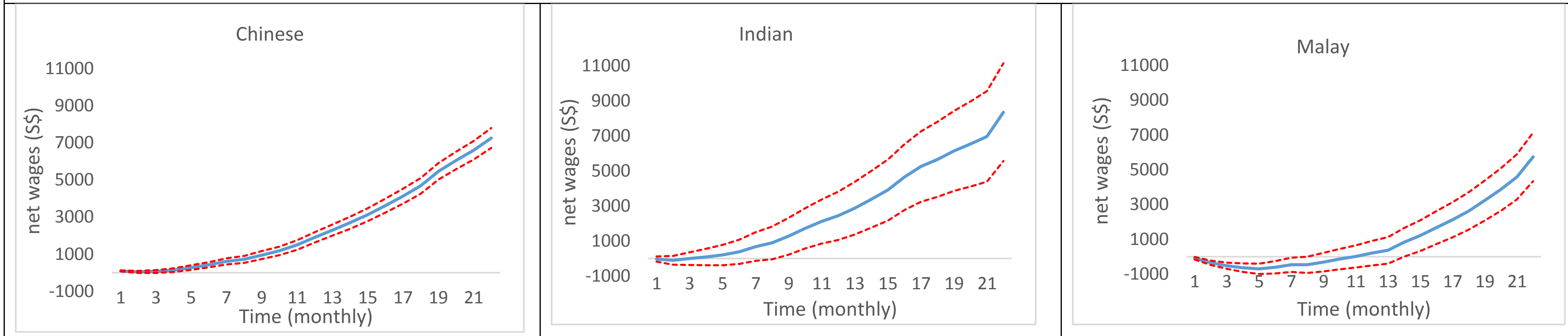
Note: The figure plots cumulative trip count (panel A) and net fares (panel B) of three groups of drivers, namely progressives, stagnant and regressive taxi drivers, using dynamic panel regressions.

Figure 8: Cumulative change of taxi trip counts & net fare of progressive drivers (only) given by race

Panel A. Dependent variable : Trip count [PC N=1,636 ; PI N= 76; PM N=162]



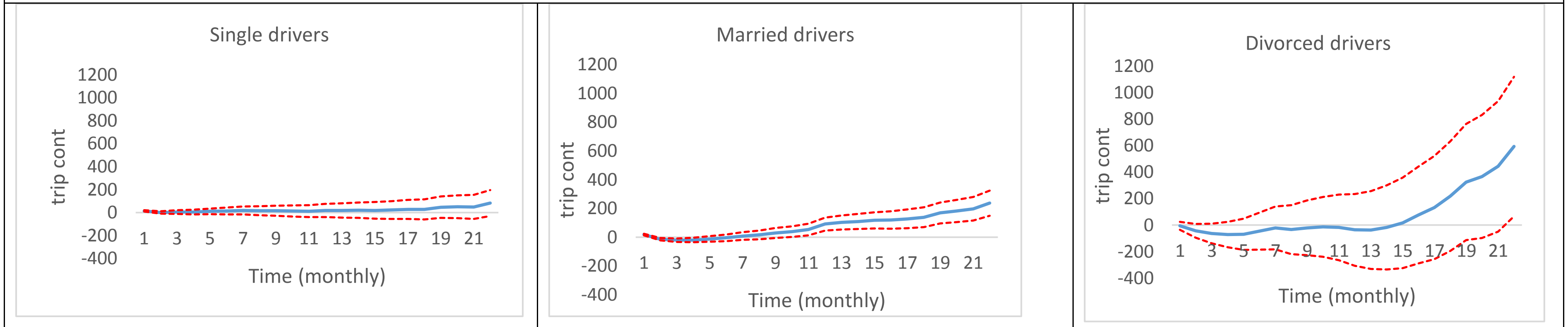
Panel B. Dependent variable : Net fare



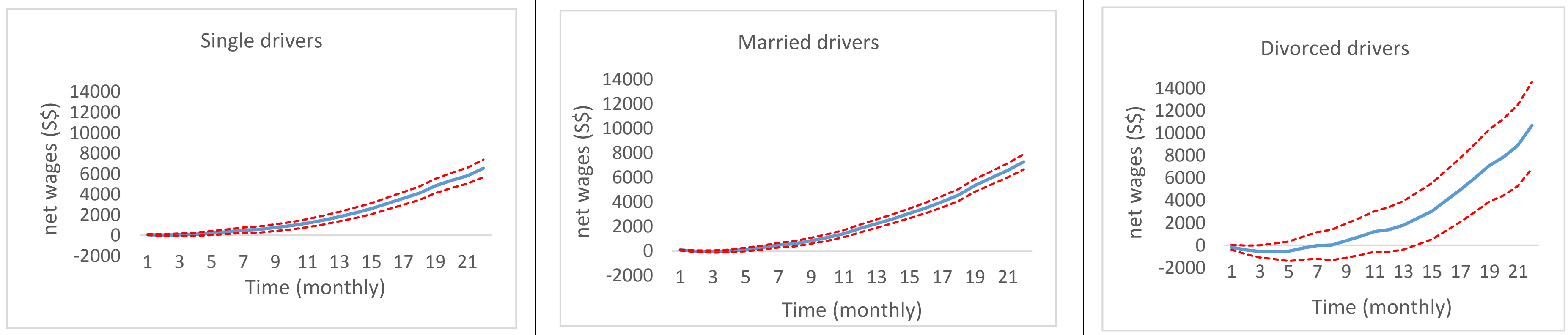
Note: The figure plots cumulative trip count (panel A) and net fares (panel B) of three groups of drivers, namely progressives, stagnant and regressive taxi drivers, using dynamic panel regressions with samples divided into different races. The sample is restricted to a total of 1,874 taxi drivers for over 23 months where 1636 are of Chinese ethnicity, 76 are of Indian ethnicity and 162 are ethnically Malay. The blue line represents the cumulative estimates (beta) of the dynamic regression along with their standard errors as red dotted lines.

Figure 9: Cumulative change of taxi trip counts & net fare of progressive drivers (only) given by marital status

Panel A. Dependent variable : Trip count [PS N= 736; PM N=1,098 PD N=34]



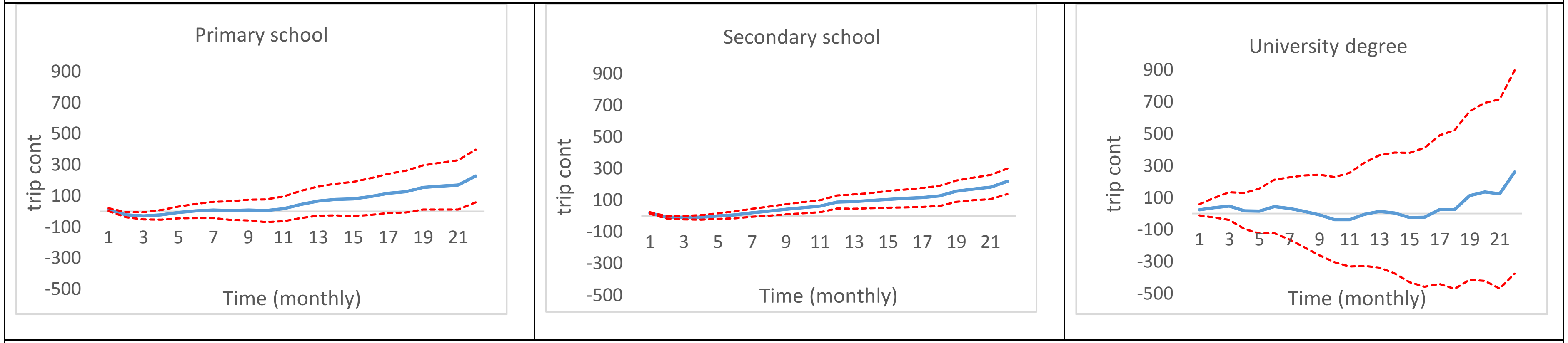
Panel B. Dependent variable : Net fare



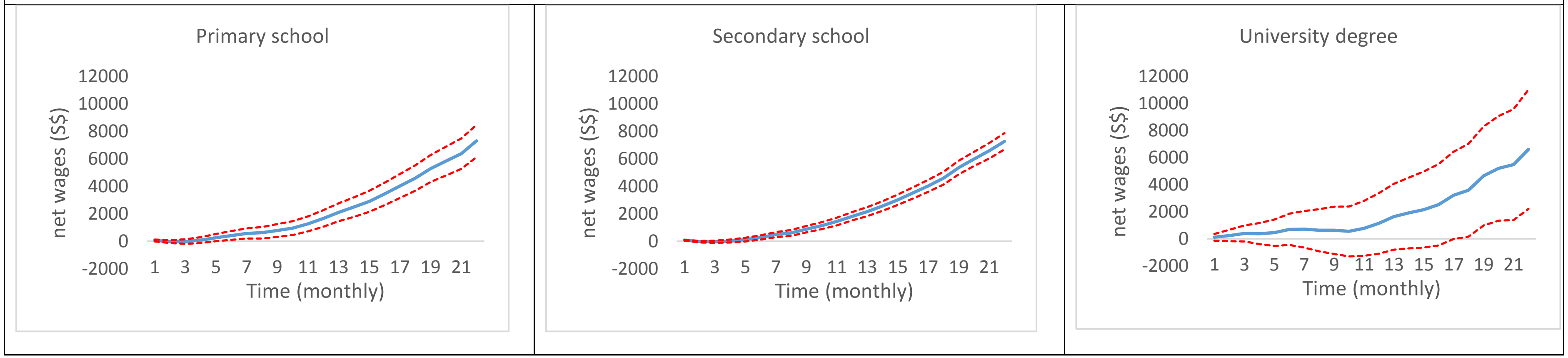
Note: The figure plots cumulative trip count (panel A) and net fares (panel B) of three groups of drivers, namely progressives, stagnant and regressive taxi drivers, using dynamic panel regressions with samples divided by marital status. The sample is restricted to a total of 1,898 taxi drivers over 23 months where 736 were single, 1098 were married, and 34 were divorced. The blue line represents the cumulative estimates (beta) of the dynamic regression along with their standard errors as red dotted lines.

Figure 10: Cumulative change of taxi trip counts & net fare of progressive drivers (only) given by education

Panel A. Dependent variable : Trip count [PP N= 318 ; PS N=1,317; PU N=35]



Panel B. Dependent variable : Net fare



Note: The figure plots cumulative trip count (panel A) and net fares (panel B) of three groups of drivers, namely progressives, stagnant and regressive taxi drivers, using dynamic panel regressions with samples divided by education level. The sample is restricted to a total of 1,670 taxi drivers over 23 months where 318 drivers hold only a primary education, 1,317 have a secondary degree, and only 35 possess a university degree. The blue line represents the cumulative estimates (beta) of the dynamic regression along with their standard errors as red dotted lines.

Tables

Table 1: OLS regression of island wide taxi driver wages against number of days driven (DD) in Singapore (2009-2010)

Dependent variable : Net wages	**OLS Model (1)**	**OLS Model (2)**	**OLS Model (3)**	**OLS Model (4)**	**OLS Model (5)**
Sample frequency	**Daily**				
Independent variables	Beta	Beta	Beta	Beta	Beta
Intercept	75.781*** (0.122)	92.868*** (1.816)	78.433*** (0.393)	84.520*** (0.198)	82.626*** (0.187)
Days driven (DD)	0.028*** (0.003)	0.027*** (0.00004)	0.026*** (0.004)	0.012*** (0.0004)	0.014*** (0.004)
Daily-Fixed Effects	NO	YES	NO	NO	NO
Monthly-Fixed Effects	NO	NO	YES	NO	NO
Yearly-Fixed Effects	NO	NO	NO	YES	NO
2009 Financial crisis dummy	NO	NO	NO	NO	YES
No of obs.	1,083,929	1,083,929	1,083,929	1,083,929	1,083,929
R square	0.004	0.096	0.007	0.007	0.006

Note: The Table (1) above reports OLS regression of days driven by a driver against his net wages for each day in Model (1). Each successive model controls for changes in demand over time by adding daily (Model (2)), monthly (Model (3)) and Yearly (Model (4)) fixed effects. The final Model (5) tests if the surge in net wages of taxi drivers in Singapore is purely due to the increase in macroeconomic activity after the financial crisis in 2009, using a dummy variable for those periods.

Table 2: Description of the variables

S.NO	Variable	Description
1	Gross wages	Total daily wages excluding daily fixed costs
2	Net wages	Net wages (gross wages-fixed costs) where fixed costs= fuel + taxi rental
4	Labor supply	Total hours spent working (daily)
5	Labor supply (daytime)	Total hours spent working (daily-daytime shift 6 a.m to 6 p.m)
6	Labor supply (nightime)	Total hours spent working (daily-nighttime shift 6 p.m to 6 a.m)
8	Trip count	Number of trip counts obtained through cruising
9	Experience	Total number of years on the driving licence
10	Age	Taxi driver age in years
11	Bottom15	Bottom 15 percentile of the wages given by two-year average
12	Middle class	Middle 70 percentile of the wages given by two-year average
13	Top15	Top 15 percentile of the wages given by two-year average
14	Trip count	Total number of daily trips
15	Labor efficiency (cruising)	Total number of daily trips obtained via booking system/ Number of hours driven (duration)
16	Labor efficiency (booking)	Total number of daily trips obtained via cruising/ Number of hours driven (duration)
17	Strategic labor supply	Number of trips taken during high demand hours of the day
18	Race	Chinese, Indian or Malay

Note: Table 2 above provides a description of all the variables and classification used in the subsequent tables. The original taxi data comes with hourly frequency which can be collapsed to daily frequency. Strategic labor supply alone is computed using hourly data that is later aggregated and matched with the daily data set.

Table 3: Descriptive statistics of key variables

year	VARIABLES	FULL SAMPLE			SOCIAL STRATIFICATION								
					bottom15			middle class			top15		
		N	Mean	StdDev	N	Mean	StdDev	N	Mean	StdDev	N	Mean	StdDev
2009	Gross wages	503135	212.81	78.2	47235	134.67	57.25	378008	207.79	68.53	77892	284.57	75.82
	Net wages	503135	75.98	74.03	47235	-0.23	54.2	378008	71.06	64.25	77892	146.04	70.99
	Labor supply	503135	775.31	214.17	47235	728.42	251.71	378008	771.03	210.68	77892	824.52	196.51
	Labor supply (daytime)	503135	496.48	184.97	47235	459.92	197.59	378008	490.33	184.16	77892	548.51	170.4
	Labor supply (nightime)	503135	278.83	164.86	47235	268.5	178.83	378008	280.7	163.92	77892	276.02	160.26
	Trip count (cruising)	503135	17.25	7.26	47235	11.34	5.57	378008	16.82	6.56	77892	22.97	7.6
	Trip count (booking)	503135	3.45	2.68	47235	1.83	1.88	378008	3.36	2.55	77892	4.85	3.03
2010	Gross wages	580794	224.21	82.99	46562	136.38	60.17	431517	216.56	72.28	102715	296.17	80.21
	Net wages	580794	88.94	78.69	46562	2.98	57.94	431517	81.68	68.02	102715	158.39	75.22
	Labor supply	580794	743.78	235.05	46562	694.69	281.5	431517	735.86	232.54	102715	799.32	211.81
	Labor supply (daytime)	580794	475.56	194.17	46562	445.18	205.95	431517	466.98	193.95	102715	525.39	180.86
	Labor supply (nightime)	580794	268.22	167.95	46562	249.51	181.95	431517	268.88	167.95	102715	273.93	160.6
	Trip count (cruising)	580794	12.47	9.57	46562	8.34	6.95	431517	12.14	9.07	102715	15.71	11.46
	Trip count (booking)	580794	8.4	8.64	46562	4.74	5.95	431517	8.02	8.17	102715	11.64	10.44

Note: *The Table (3) above provides the descriptive statistics of wages and labor supply related variables subsequent tables. The sample is split both temporally (2009 & 2010) and by social stratification such as* bottom15, top15, *and* middle class *which comprises the remaining 70 percent of drivers in the middle. The* 2yr avg *in the variable column represents the two-year average of the variables concerned.*

Table 4: Descriptive statistics of key variables by race

		FULL SAMPLE			2009			2010		
Ethnicity	**Variables**	**N**	**Mean**	**StdDev**	**N**	**Mean**	**StdDev**	**N**	**Mean**	**StdDev**
CHINESE	Gross wages	948389	220.06	81.2	445464	213.83	78.4	502925	225.58	83.22
	Net wages	948389	84.15	77.03	445464	77.04	74.26	502925	90.45	78.86
	Labor supply	948389	757.8	223.06	445464	774.89	211.25	502925	742.67	231.97
	Labor supply (daytime)	948389	487.46	189.71	445464	498.47	184.54	502925	477.71	193.66
	Labor supply (nightime)	948389	270.34	165.18	445464	276.42	163.52	502925	264.96	166.45
	Trip count (cruising)	948389	14.88	8.93	445464	17.44	7.26	502925	12.62	9.63
	Trip count (booking)	948389	6.1	7.06	445464	3.44	2.68	502925	8.45	8.71
INDIAN	Gross wages	34801	231.62	84.71	14901	226.15	82.36	19900	235.71	86.2
	Net wages	34801	95.01	80.19	14901	88.5	77.78	19900	99.88	81.62
	Labor supply	34801	775.05	231.43	14901	786.45	215.57	19900	766.51	242.29
	Labor supply (daytime)	34801	480.8	185.17	14901	487.74	179.12	19900	475.6	189.4
	Labor supply (nightime)	34801	294.25	159.67	14901	298.71	160.63	19900	290.91	158.88
	Trip count (cruising)	34801	14.26	8.95	14901	16.89	7.51	19900	12.29	9.42
	Trip count (booking)	34801	6.66	6.9	14901	4	2.83	19900	8.65	8.24
MALAY	Gross wages	100739	203.78	75.8	42770	197.48	72.45	57969	208.43	77.85
	Net wages	100739	67.18	71.61	42770	60.53	68	57969	72.08	73.79
	Labor supply	100739	758.41	251.51	42770	775.76	241.98	57969	745.61	257.57
	Labor supply (daytime)	100739	466.19	195.75	42770	478.77	190.35	57969	456.91	199.13
	Labor supply (nightime)	100739	292.22	180.01	42770	296.99	178.1	57969	288.7	181.32
	Trip count (cruising)	100739	13.03	8.42	42770	15.46	6.89	57969	11.24	8.97
	Trip count (booking)	100739	5.94	6.82	42770	3.32	2.65	57969	7.86	8.19

Note: *The Table (4) above provides the descriptive statistics of wages and labor supply related variables when drivers are divided by ethnicity' in our sample, namely Chinese, Indian and Malays . Again the sample is split both temporally (2009 & 2010) as in Table (2) to give yearly progression of the variables concerned. The* 2yr avg *in the variable column represents the two-year average of the variables concerned.*

Table 5 (a): Descriptive statistics given by progressive group as classified using AR(1)

GROUP		FULL SAMPLE			social stratification								
					bottom15			*middle class*			*top15*		
	Variables	N	Mean	StdDev	N	Mean	StdDev	N	Mean	StdDev	N	Mean	StdDev
Progressive group	Gross wages	737473	218.46	80.92	68172	135.78	59.46	554234	213.26	70.89	115067	292.5	77.49
	Net wages	737473	82.02	77.14	68172	0.75	57.41	554234	76.93	66.9	115067	154.69	72.86
	Labor supply	737473	765.61	225.03	68172	718.97	264.79	554234	761.12	221.81	115067	814.83	205.5
	Labor supply (daytime)	737473	489.35	164.48	68172	459.76	177.98	554234	484.16	163.17	115067	531.9	161.71
	Labor supply (nightime)	737473	276.25	134.2	68172	259.21	169.24	554234	276.96	127.86	115067	282.93	125.78
	Trip count (cruising)	737473	14.8	8.88	68172	9.72	6.54	554234	14.46	8.32	115067	19.42	10.47
	Trip count (booking)	737473	5.97	6.93	68172	3.31	4.68	554234	5.78	6.6	115067	8.45	8.65

Note: *Table 5 (a) above provides the descriptive statistics of wages and labor supply related variables for taxi drivers with increasing wages by AR(1) model, alternatively called the progressive group . The sample spans two years (2009 & 2010) excluding the month of December in 2009. The descriptive statistics are given by different social stratifications using the* net wages *variable, drivers are split into the* bottom15 *percent, the*top15 *percent and* middle class *which comprises the remaining 70 percent in the middle. Total number of taxi drivers in the sample = 1,874.*

Table 5 (b): Descriptive statistics given by regressive group as classified using AR(1)

	Variables	**FULL SAMPLE**			**social stratification**								
					bottom15			*middle class*			*top15*		
		N	Mean	StdDev	N	Mean	StdDev	N	Mean	StdDev	N	Mean	StdDev
Regressive group	Gross wages	26941	225	84.8	1859	117.64	58.99	21252	220.54	75.11	3830	301.83	77.37
	Net wages	26941	91.3	78.15	1859	4.27	51.4	21252	85.83	70.03	3830	163.93	73.16
	Labor supply	26941	744.78	241.13	1859	495.23	211.37	21252	755.29	241.92	3830	807.64	166.75
	Labor supply (daytime)	26941	435.17	175.74	1859	258.24	127.38	21252	430.86	180.16	3830	544.96	152.64
	Labor supply (nightime)	26941	309.62	146.83	1859	236.99	122.04	21252	324.43	137.41	3830	262.68	61.03
	Experience	26941	29.64	7.23	1859	31.05	2.61	21252	29.65	7.47	3830	28.92	7.3
	Trip count (cruising)	26941	15.14	9.09	1859	9.3	6.58	21252	14.97	8.6	3830	18.93	10.9
	Trip count (booking)	26941	6.27	7.16	1859	2.7	4.8	21252	6.04	6.81	3830	9.25	8.73

Note: *Table 5 (b) above provides the descriptive statistics of wages and labor supply related variables for taxi drivers with decreasing wages identified by AR(1) model, alternatively called the regressive group . The sample spans two years (2009 & 2010) excluding the month of December in 2009. The descriptive statistics are given by different social stratifications using the* net wage' *variable, drivers are split into the* bottom15 *percent, the*top1' *percent and* middle class *which comprises the remaining 70 percent in the middle. Total number of taxi drivers in the sample = 107.*

Table 5 (c): Descriptive statistics given by stagnant group as classified using AR(1)

GROUP	Variables	FULL SAMPLE			social stratification								
					bottom15			*middle class*			*top15*		
		N	Mean	StdDev	N	Mean	StdDev	N	Mean	StdDev	N	Mean	StdDev
Stagnant group	Gross wages	319515	219.46	80.85	23766	136.18	56.3	234039	209.84	69.71	61710	288.02	80.47
	Net wages	319515	84.3	75.95	23766	2.89	52.53	234039	75.41	65.12	61710	149.37	75.05
	Labor supply	319515	742.97	226.58	23766	707.7	271.76	234039	731.06	223.6	61710	801.7	208.05
	Labor supply (daytime)	319515	480.08	169.95	23766	447.27	191.33	234039	467.28	170.8	61710	541.22	157.5
	Labor supply (nightime)	319515	262.89	136.98	23766	260.42	186.93	234039	263.78	132.76	61710	260.47	111.35
	Experience	319515	28.91	9.38	23766	31.44	10.17	234039	28.96	9.51	61710	27.74	8.32
	Trip count (cruising)	319515	14.41	8.92	23766	10.26	6.26	234039	13.95	8.32	61710	17.77	10.76
	Trip count (booking)	319515	6.39	7.25	23766	3.21	4.49	234039	5.97	6.7	61710	9.17	9.02

Note: *Table 5 (c) above provides the descriptive statistics of wages and labor supply related variables for taxi drivers with stagnant wages identified by AR(1) model, alternatively called the stagnant' group . The sample spans two years (2009 & 2010) excluding the month of December in 2009. The descriptive statistics are given by different social stratifications using the* net wages *variable, drivers are split into the* bottom15 *percent, the* top1' *percent, and* middle class *which comprises the remaining 70 percent in the middle. Total number of taxi drivers in the sample = 1,835.*

Table 6: Determinants of daily wages among Singaporean taxi drivers estimated using a standard OLS regression

Dependent variable: ***Net wages***	**Model (1) OLS**		**Model (2) OLS**	
Independent variables	Beta	Std. error	Beta	Std. error
Intercept	-266.043***	2.2605	-244.297***	2.501
Labor supply	32.008***	0.062	30.170***	0.062
Labor supply sq	-0.803***	0.002	-0.758***	0.002
Labor efficiency B (booking)	39.266***	0.086	37.175***	0.113
Labor efficiency C (cruising)	29.461***	0.068	28.585***	0.068
Strategic labor supply	55.103	0.139	53.426***	0.139
age	-0.023	0.086	0.261***	0.082
agesq	-0.007***	0.000	-0.010***	0.000
Chinese (base=Malay)	3.431***	0.188	4.360***	0.181
Indian (base=Malay)	10.782***	0.332	11.338***	0.320
Marital status	-0.132***	0.111	-0.037	0.107
Secondary (base=Primary)	0.391***	0.143	0.440***	0.138
Diploma (base=Primary)	-2.374***	0.478	-2.540***	0.460
Pupil Referral Unit (PRU) (base=Primary)	4.304***	0.287	4.378***	0.276
Polytechnic (POLY) (base=Primary)	4.371	0.376	4.6638**	0.362
Institute of Technical Education (ITE) (base=Primary)	-7.495	0.338	-7.775***	0.326
University (UNI) (base=Primary)	-2.503	0.424	-2.040***	0.409
DAY FIXED EFFECTS	NO		YES	
MONTHLY FIXED EFFECTS	NO		YES	
YEAR FIXED EFFECTS	NO		YES	
No of obs.	1,083,929		1,083,929	
R square	0.79		0.81	

Note: *The table reports the regression results from the panel regressions estimated using Eq. (4) with 'net wages'(S$) as the dependent variable with Model (2) controlling for day, monthly and year fixed effects.* Labor supply/Labor supply sq *are the total labor supply (in hrs) given and its square.* Labor efficiency B/C *are all taxi trips obtained through automated booking system/street cruising in Singapore normalized by labor supply. Other control variables include age of the driver and its square (*age *and* agesq*) and two dummy variables that identify race of taxi drivers (*Chinese, Indian *and* Malay*), where* Malay *is kept as base dummy. A dummy variable is use for marital status where 0-unmarried and 1-married (*Marital status*). In addition, six dummy variables are included to identify educational qualifications of taxi drivers (*Primary, Secondary, *Diploma, PRU, POLY, ITE, and UNI), where primary education is kept as base dummy.* Strategic labor supply *represents all passenger trips that are taken during peak hours normalized by labor supply, these usually include few hours in the morning and evening in Singapore. The standard errors are shown with coefficients on the right. *** indicates significance at the 1% level; ** indicates significance at the 5% level; * indicates significance at the 10% level.*

Table 7 (a): The impact of learning on daily wages of Singaporean taxi drivers given by groups

Dependent variable : Log(Net wages)	**Model (1) OLS**		**Model (2) OLS**		**Model (1) OLS**		**Model (2) OLS**		**Model (1) OLS**		**Model (2) OLS**	
Independent variables	**Beta**	**Std. error**	**Beta**	**Std. error**	**Beta**	**Std. error**	**Beta**	**Std. error**	**Beta**	**Std. error**	**Beta**	**Std. error**
Sample	**Progressive**				**Stagnant**				**Regressive**			
Intercept	4.117***	0.012	4.117***	0.012	4.485***	0.015	4.485***	0.015	3.916***	0.059	3.916***	0.059
Labor supply	0.126***	0.000	0.126***	0.000	0.124***	0.000	0.124***	0.000	0.132***	0.001	0.132***	0.001
Labor supply sq	-0.003***	0.000	-0.003***	0.000	-0.003***	0.000	-0.003***	0.000	-0.003***	0.000	-0.003***	0.000
labor efficiency B (booking)	0.160***	0.000	0.160***	0.000	0.108***	0.000	0.108***	0.000	0.095***	0.002	0.095***	0.002
labor efficiency C (cruising)	0.126***	0.000	0.126***	0.000	0.089***	0.000	0.089***	0.000	0.055***	0.001	0.055***	0.001
Strategic labor supply	0.208***	0.000	0.208***	0.000	0.209***	0.000	0.209***	0.000	0.267***	0.003	0.267***	0.003
age	0.005***	0.000	0.005***	0.000	-0.005***	0.000	-0.005***	0.000	0.012***	0.001	0.012***	0.001
Agesq	-0.000***	0.000	-0.000***	0.000	0.000***	0.000	0.000***	0.000	-0.000***	0.000	-0.000***	0.000
Chinese (base=Malay)	0.005***	0.000	0.005***	0.000	0.015***	0.001	0.015***	0.001	0.040***	0.003	0.040***	0.003
Indian (base=Malay)	0.035***	0.001	0.035***	0.001	0.018***	0.002	0.018***	0.002	-0.043***	0.014	-0.043***	0.014
Marital status	-0.004***	0.000	-0.004***	0.000	0.001*	0.000	0.001*	0.000	0.010***	0.002	0.010***	0.002
Secondary	-0.004***	0.000	-0.004***	0.000	0.011***	0.000	0.011***	0.000	0.037***	0.003	0.037***	0.003
Diploma	-0.014***	0.002	-0.014***	0.002	-0.009***	0.002	-0.009***	0.002	0.017***	0.005	0.017***	0.005
PRU	0.001	0.001	0.001	0.001	0.031***	0.002	0.031***	0.002	0.001	0.005	0.001	0.005
POLY	-0.001	0.001	-0.001	0.001	0.042***	0.002	0.042***	0.002	0.046**	0.018	0.0463**	0.018
ITE	-0.030***	0.001	-0.030***	0.001	-0.002	0.002	-0.002	0.002	-0.024**	0.011	-0.024**	0.011
UNI	0.005***	0.001	0.005***	0.001	-0.051***	0.003	-0.051***	0.003	-	-	-	-
DAY FIXED EFFECTS	YES		YES		YES		YES		YES		YES	
MONTHLY FIXED EFFECTS	NO		YES		NO		YES		NO		YES	
YEAR FIXED EFFECTS	NO		YES		NO		YES		NO		YES	
No of obs.	737,473		737,473		319,515		319,515		26,941		26,941	
R square	0.59		0.61		0.60		0.60		0.67		0.67	

Note: *The table reports the regression results from the panel regressions estimated using Eq. (4) with* net wages*(S$) as the dependent variable with Models (1) and (2) controlling for day (1, 2), monthly and year (2) fixed effects respectively.* Labor supply/Labor supply sq are *the total labor supply (in hrs) given and its square.* Labor efficiency B/C *are all taxi trips obtained through automated booking system/street cruising in Singapore normalized by labor supply. Other control variables include age of the driver and its square* (age, agesq) *and two dummy variables that identify race of taxi drivers (*Chinese, Indian *and* Malay), *where* Malay *is kept as base dummy. A dummy variable is used for marital status where 0-unmarried and 1-married (*Marital status*). In addition, six dummy variables are included to identify educational qualifications of taxi drivers* (Primary, Secondary, Diploma, PRU, POLY, ITE, and UNI*), where primary education is kept as base dummy*. Strategic labor supply *represents all passenger trips that are taken during peak hours normalized by labor supply, these usually include few hours in the morning and evening in Singapore. The standard errors are shown with coefficients on the right. *** indicates significance at the 1% level; ** indicates significance at the 5% level; * indicates significance at the 10% level.*

Table 7 (b): The impact of learning on daily wages of Singaporean taxi drivers given by groups

Dependent variable : Log(Net wages)	**Model (2) OLS**		**Model (2) OLS**		**Model (2) OLS**	
Independent variables	**Beta**	**Std. error**	**Beta**	**Std. error**	**Beta**	**Std. error**
Sample	**Progressive**		**Stagnant**		**Regressive**	
Intercept	4.121***	0.021	4.326***	0.012	4.409***	0.085
Labor supply	0.128***	0.000	0.124***	0.000	0.129***	0.001
Labor supply_sq	-0.003***	0.000	-0.003***	0.000	-0.003***	0.000
Labor efficiency (booking)	0.127***	0.000	0.089***	0.000	0.085***	0.002
Labor efficiency (cruising)	0.111***	0.000	0.083***	0.000	0.048***	0.001
Strategic labor supply	0.229***	0.000	0.208***	0.000	0.261***	0.003
DAY FIXED EFFECTS	YES		YES		YES	
DRIVER FIXED EFFECTS	YES		YES		YES	
No of obs.	737,473		319,515		26,941	
R square	0.68		0.68		0.71	

Note: *The table reports the regression results from the panel regressions estimated using Eq. (5) with log of* net wages *as the dependent variable.* Labor supply *is the total labor supply (in hrs) given by taxi whereas* labor efficiency *is taxi trips obtained through cruising/automated booking system normalised by labor supply (hrs) in Singapore*. Strategic labor supply *represents all passenger trips that are taken during peak hours normalized by labor supply, these usually include few hours in the morning and evening in Singapore. The standard errors are shown with coefficients on the right. *** indicates significance at the 1% level; ** indicates significance at the 5% level; * indicates significance at the 10% level.*

Table 8 (a): The impact of learning on daily wages of Singaporean taxi drivers - sample split by year

Years	2009						2010					
	Progressive group		Stagnant group		Regressive group		Progressive group		Stagnant group		Regressive group	
Dependent variable : Log(Net wages)	Model (1) OLS		Model (1) OLS		Model (1) OLS		Model (1) OLS		Model (1) OLS		Model (1) OLS	
Independent variables	Beta	Std. error	Beta	Std. error	Beta	Std. error	Beta	Std. error	Beta	Std. error	Beta	Std. error
Intercept	4.060***	0.015	4.424***	0.022	4.459***	0.085	4.158***	0.015	4.460***	0.018	3.713***	0.072
Labor supply	0.115***	0.000	0.112***	0.000	0.119***	0.001	0.130***	0.000	0.129***	0.000	0.136***	0.001
Labor supply sq	-0.002***	0.000	-0.002***	0.000	-0.003***	0.000	-0.003***	0.000	-0.003***	0.000	-0.003***	0.000
Labor efficiency B (booking)	0.464***	0.001	0.384***	0.002	0.357***	0.009	0.134***	0.000	0.096***	0.000	0.086***	0.002
Labor efficiency C (cruising)	0.144***	0.000	0.082***	0.000	0.047***	0.001	0.117***	0.000	0.102***	0.000	0.073***	0.002
Strategic labor supply	0.196***	0.000	0.211***	0.001	0.288***	0.004	0.203***	0.000	0.195***	0.001	0.237***	0.004
age	0.004***	0.000	-0.005***	0.000	-0.012***	0.003	0.004***	0.000	-0.004***	0.000	0.021***	0.002
Agesq	-0.000***	0.000	0.000***	0.000	0.000***	0.000	-0.000***	0.000	0.000***	0.000	-0.000***	0.000
Chinese (base=Malay)	0.006***	0.001	0.004***	0.001	0.032***	0.006	0.006***	0.001	0.023***	0.001	0.038***	0.004
Indian (base=Malay)	0.034***	0.002	0.002	0.003	-0.000	0.014	0.032***	0.002	0.023***	0.002	0.230***	0.086
Marital status	-0.002***	0.000	0.008***	0.001	0.000	0.003	-0.004***	0.000	-0.003***	0.000	0.011***	0.003
Secondary	-0.016***	0.000	0.004***	0.001	0.026***	0.004	0.000	0.000	0.014***	0.001	0.041***	0.004
Diploma	-0.018***	0.003	-0.030***	0.004	0.015*	0.009	-0.018***	0.003	0.002	0.003	0.021***	0.007
PRU	-0.012***	0.001	0.011***	0.003	-0.028***	0.007	0.006***	0.001	0.035***	0.002	0.021***	0.007
POLY	-0.022***	0.002	0.017***	0.003	0.079***	0.021	0.007***	0.002	0.045***	0.003	0.067**	0.032
ITE	-0.032***	0.002	0.012***	0.003	-	-	-0.030***	0.002	-0.005*	0.002	-	-
UNI	-0.004*	0.002	-0.065***	0.004	-	-	0.013***	0.002	-0.025***	0.004	-	-
DAY FIXED EFFECTS	YES		YES		YES		YES		YES		YES	
MONTHLY FIXED EFFECTS	YES		YES		YES		YES		YES		YES	
No of obs.	345,188		145,535		12,412		392,285		173,980		14,529	
R square	0.65		0.63		0.72		0.58		0.60		0.66	

Note: *The table reports the regression results from the panel regressions estimated using Eq. (5) with log of* net wages *as the dependent variable.* Labor supply *is the total labor supply (in hrs) given by taxi whereas* labor efficiency *B/C are taxi trips obtained through automated booking system/cruising normalized by labor supply (hrs) in Singapore. Other control variables include age of the driver and its square (*age *and* agesq*), number of years on driving license (*experience*), and two dummy variables that identify ethnicities of taxi drivers (*Chinese, Indian *and* Malay*) where* Malay *is kept as base. A dummy variable is used for marital status where 0-unmarried and 1-married (*Marital status*). In addition, six dummy variables are included to identify educational qualifications of taxi drivers (Primary, Secondary, Diploma, PRU, POLY, ITE, and UNI) where drivers with primary education are the base dummy.* Strategic labor supply *represents all passenger trips that are taken during peak hours normalized by labor supply, these usually include few hours in the morning and evening in Singapore. The standard errors are shown with coefficients on the right. *** indicates significance at the 1% level; ** indicates significance at the 5% level; * indicates significance at the 10% level.*

Table 8 (b): The impact of learning on daily wages of Singaporean taxi drivers - sample split by year

Years	2009						2010					
	Progressive group		Stagnant group		Regressive group		Progressive group		Stagnant group		Regressive group	
Dependent variable : Log(Net wages)	Model (2) OLS		Model (2) OLS		Model (2) OLS		Model (2) OLS		Model (2) OLS		Model (2) OLS	
Independent variables	Beta	Std. error	Beta	Std. error	Beta	Std. error	Beta	Std. error	Beta	Std. error	Beta	Std. error
Intercept	4.115***	0.020	4.245***	0.016	4.185***	0.030	4.246***	0.030	4.353***	0.016	4.421***	0.087
Labor supply	0.117***	0.000	0.113***	0.001	0.115***	0.002	0.132***	0.000	0.128***	0.001	0.134***	0.002
Labor supply sq	-0.003***	0.000	-0.003***	0.000	-0.003***	0.000	-0.004***	0.000	-0.003***	0.000	-0.004***	0.000
Labor efficiency B (booking)	0.424***	0.002	0.352***	0.003	0.380***	0.010	0.109***	0.001	0.081***	0.001	0.077***	0.002
Labor efficiency C (cruising)	0.118***	0.000	0.070***	0.001	0.042***	0.001	0.107***	0.000	0.101***	0.001	0.064***	0.002
Strategic labor supply	0.224***	0.001	0.209***	0.001	0.265***	0.005	0.219***	0.001	0.197***	0.001	0.241***	0.004
DAY FIXED EFFECTS	YES		YES		YES		YES		YES		YES	
DRIVER FIXED EFFECTS	YES		YES		YES		YES		YES		YES	
No of obs.	345,188		145,535		12,412		392,285		173,980		14,529	
R square	0.68		0.68		0.76		0.63		0.66		0.71	

Note: *The table reports the regression results from the panel regressions estimated using Eq. (5) with log of* net wages *as the dependent variable.* Labor supply *is the total labor supply (in hrs) given by taxi whereas* labor efficiency B/C *are taxi trips obtained through automated booking system/cruising normalized by labor supply (hrs) in Singapore.* Strategic labor supply *represents all passenger trips that are taken during peak hours normalized by labor supply, these usually include few hours in the morning and evening in Singapore. The standard errors are shown with coefficients on the right. *** indicates significance at the 1% level; ** indicates significance at the 5% level; * indicates significance at the 10% level.*