

天气冲击、动态调价与劳动供给

——来自上海出租车市场的证据

马艺瑗 曹思力 朱喜 朱平芳

目 录

附录 I 原始行程数据介绍	1
附录 II 行程数据结构化为小时数据	2
附录 III 行程数据结构化为班次数据	3
附录 IV 主要变量的描述性统计	5
附录 V 降雨的分布情况	7
附录 VI 临时休息决策	9
附录 VII 稳健性分析一：区分全职和兼职驾驶员	10
附录 VIII 稳健性分析二：高峰时段动态调价的影响	12
附录 IX 稳健性分析三：不同时间段的异质性分析	13

附录 I 原始行程数据介绍

本文出租车驾驶员数据来自上海市出租车市场，每笔行程为驾驶员的一次载客记录如表 II 所示，该记录包含驾驶员编号、车牌编号、上车时间、下车时间、行驶里程、营运收入等信息。

表 II 行程的主要信息

行程编号	驾驶员编号	车牌编号	上车时间	下车时间	行驶里程 (公里)	营运收入 (元)
1	2505	106	2015/1/1 5:37	2015/1/1 5:49	9	28
2	2505	106	2015/1/1 6:01	2015/1/1 6:11	8	26

原始数据为出租车乘客出行的行程信息，并非直接与驾驶员的劳动供给概念相对应，本文参照经典文献的做法，在实证模型中选择了两种对劳动供给的刻画方式：

一是小时为单位的劳动供给，即某小时内是否在出租车市场提供有效劳动，具体构造方式见附录 II。

二是以班次为单位的劳动供给，即驾驶员在某天出车开始工作到收车结束工作的整个过程，重点关注结束工作的决策，具体构造方式见附录 III。

附录 II 行程数据结构化为小时数据

对于单个驾驶员而言，考察其每一笔行程的起止时间与每个小时单位的交集，如图 III1 所示，与 $T:00-(T+1):00$ 小时的交集即为图中阴影部分。小时载客时长为该小时内驾驶员处于运营载客状态的时间长度总和，即阴影部分时长的加总。

小时收入则根据以下规则计算：按照行程在该小时中的时长占行程总时长的比例，将行程收入折算到对应小时当中；然后，对小时内所有交集行程的折算收入进行加总。小时载客里程按同样的方法进行计算。小时载客速度由小时载客里程除以小时载客时长得到。

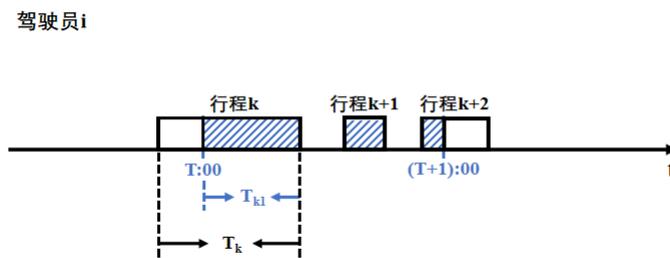


图 III1 行程信息结构化为小时信息示意图

此外，将小时内所有工作的驾驶员的小时信息进行平均，可以得到在某个小时工作驾驶员的平均小时收入、小时载客时间、小时载客里程。将小时内有载客记录的驾驶员数量加总，可以得到每小时的工作人数。

附录 III 行程数据结构化为班次数据

虽然驾驶员无法决定每一笔行程的持续时间和里程,但是可以决定每一次何时何地上班以及何时何地下班,将驾驶员从出车上班到收车下班回家的一整个劳动供给过程定义为“班次”概念。由于驾驶员原始行程的数据集仅按照时间顺序对载客记录进行排序,并不包含上下班时刻的节点信息,因而这一节点需要按照一定标准进行识别。又由于数据是按照驾驶员进行随机抽样,每一位驾驶员的行程信息在抽样的时间范围内保证了完整性,因而驾驶员的上下班节点就包含在某些行程之间。

考虑到驾驶员载客行程之间的间隔可能源于正常的乘客搜寻、短暂休息或者收车回家,并且收车回家的时间比前两者要长得多,因此需要找到一个合适的标准进行最大程度的区分。由于 90% 以上的行程间隔都小于 2 小时,考虑大于 2 小时、小于 24 小时的行程间隔时长,以 2019 年数据集内的网约车和巡游出租车为例,行程之间间隔时长的分布如图 III1 所示。

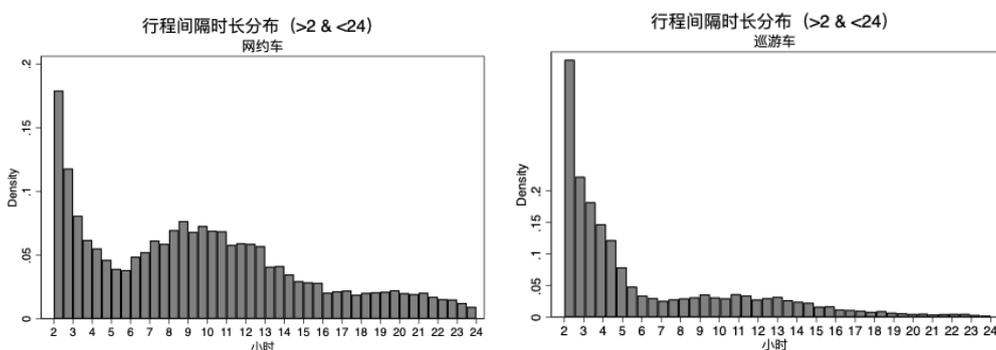


图 III1 网约车（左）和巡游车（右）行程之间 (>2h) 间隔时长分布

从行程时间间隔的分布上可以看出,网约车和巡游车驾驶员的行程间隔时长分布分别在 5 小时和 6 小时处呈现出一个低谷。因此,分别以 5 小时和 6 小时为界,可以将乘客搜寻、短暂休息的行为,与收车下班的行为进行最大程度地区分。

如图 III2 所示,以巡游车为例,本文以 6 小时为界,将驾驶员收车下班视为一个班次的结束,将行程之间间隔时长小于 6 小时的情况视为一个班次内的正常搜寻/短暂休息;将某行程与下一个行程之间间隔时长大于 6 小时的情况视为一个班次的结束,下一个行程视为一个新的班次的开始。

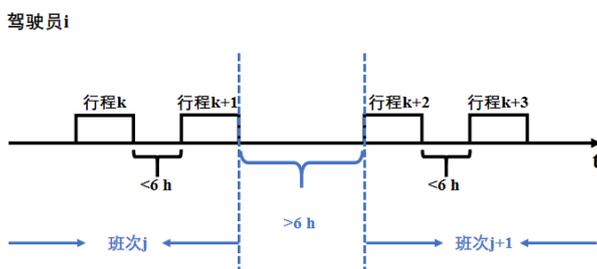


图 III2 驾驶员班次划分标准（以巡游车为例）

每个班次结束时间与开始时间之差为班次总时长,将班次内所有行程的营运收入和载客里程进行加总,可以得到班次的营运收入和载客里程信息。此外,以班次中每一个行程结束

作为节点，每个行程结束时间与班次开始时间之差为累积工作时长，将班次内已经完成行程的营运收入进行加总，可以得到累积营运收入。

网约车按照同样的方法进行转化，主要差别在于，根据行程间隔分布的特点，将 6 小时的划分标准变为 5 小时。

附录 IV 主要变量的描述性统计

本文实证部分所使用的 2015 年—2017 年巡游驾驶员、2019 年巡游车和网约车驾驶员样本, 转化为小时信息后的变量描述性统计如表 IV1 所示。可以看出, 三组驾驶员平均的小时载客时长、营运收入和载客里程较为接近, 尤其是 2019 年同期的巡游车和网约车, 具有较好的可比性。值得注意的是, 网约车虽然小时营运收入的平均值略低于巡游车, 但标准差高于巡游车, 这反映了网约车运价机制的灵活性。

表 IV1 驾驶员小时信息描述性统计

时段	驾驶员类型	变量	样本容量	均值	标准差	最小值	最大值
2015 年—2017 年	巡游车	营运收入(元)	20,452,430	62.986	33.091	0.013	318
		载客时长(小时)	20,452,430	0.527	0.242	0.017	1
		载客里程(公里)	20,452,430	14.727	8.996	0	75.246
2019 年	巡游车	营运收入(元)	220,111	59.552	34.561	0.014	311.4752
		载客时长(小时)	220,111	0.487	0.245	0.000	1
		载客里程(公里)	220,111	14.011	9.334	0.003	75.069
2019 年	网约车	营运收入(元)	257,821	55.134	38.951	0.009	549.830
		载客时长(小时)	257,821	0.511	0.264	0.000	1
		载客里程(公里)	257,821	11.803	8.480	0	75.622

2019 年巡游车和网约车驾驶员样本转化为班次信息后的变量描述性统计如表 IV2 所示。可以看出, 巡游车班次的时长、营收和里程普遍高于网约车, 班次数量显著低于网约车。这主要是由于巡游车主要为全职驾驶员, 工作过程较为连贯; 而网约车驾驶员中有一定的兼职驾驶员, 通常在早晚高峰临时补充到出租车市场中, 这类驾驶员普遍班次时间较短、班次数量较多。基于这一事实, 本文将在稳健性分析部分对网约车驾驶员进行全职和兼职类型的区分。

表 IV2 驾驶员班次信息描述性统计

时段	驾驶员类型	变量	样本容量	均值	标准差	最小值	最大值
2019 年	巡游车	营运收入(元)	13,766	823.909	371.289	19	2346
		班次时长(小时)	13,766	14.495	5.544	0.291	33.096
		载客里程(公里)	13,766	194.500	89.615	1	607.900
2019 年	网约车	营运收入(元)	27,567	459.594	336.174	9.680	2413.570
		班次时长(小时)	27,567	7.802	5.327	0.146	35.499
		载客里程(公里)	27,567	98.661	70.253	0	446.900

降雨是本文的核心解释变量, 2015 年—2017 年以及 2019 年某一周内降雨记录的小时降雨量(单位为 mm)描述性统计如表 IV3 所示。2015 年—2017 年, 26 304 个小时中有 2 937 个小时的降雨量大于 0。2019 年一周的跨度较短, 168 个小时中有 39 个小时的降雨量大于 0, 较好地支撑了本文的实证研究。

表 IV3 降雨描述性统计

时段	变量	样本容量	均值	标准差	最小值	最大值
2015 年—2017 年	降雨量 (mm)	2,937	1.578	3.139	0.100	65.400
2019 年	降雨量 (mm)	39	3.610	5.521	0.100	27.600

附录 V 降雨的分布情况

本文样本所在的时期包含下雨和非下雨的时段。其中，2015—2017 年样本时间跨度较长，包含丰富的降雨信息；2019 年一周样本时间跨度较短，因此，本节对这一周的降雨分布情况进行讨论。

2019 年一周样本中，降雨在每天的分布情况如表 V1 所示，降雨记录均出现在工作日，未出现在休息日。由于工作日和休息日居民出行模式不同，且本文样本中休息日并未出现降雨，如果将休息日纳入模型当中，降雨对劳动供给的影响会被工作日与休息日的区别所干扰。因此，本文在利用 2019 年一周数据进行的劳动供给的实证研究中不考虑休息日。

表 V1 2019 年样本各天降雨的分布

星期	不下雨的小时数	下雨的小时数	总计
星期日	24	0	24
星期一	16	8	24
星期二	9	15	24
星期三	24	0	24
星期四	10	14	24
星期五	22	2	24
星期六	24	0	24
总计	129	39	168

此外，出租车市场有着明显的周期性，工作日主要由通勤的周期决定。本文根据出租车驾驶员载客的低速行驶时间段，将工作日早晨的 7:00—8:00、8:00—9:00、9:00—10:00，以及晚上的 17:00—18:00、18:00—19:00，分别界定为早、晚高峰时段。进一步地，将工作日的一天划分为如表 V2 所示几个时段，其中第（2）列表示早、晚高峰的两个时段，这两个时段又将一天的其他 19 个小时划分为另外三个部分，第（1）列为凌晨时段，第（3）列为早、晚高峰之间的日间平峰时段，第（4）列为晚高峰之后的夜间到午夜之间的时段。

表 V2 工作日小时的分组

(1)	(2)	(3)	(4)
0:00—1:00	7:00—8:00	10:00—11:00	19:00—20:00
1:00—2:00	8:00—9:00	11:00—12:00	20:00—21:00
2:00—3:00	9:00—10:00	12:00—13:00	21:00—22:00
3:00—4:00	17:00—18:00	13:00—14:00	22:00—23:00
4:00—5:00	18:00—19:00	14:00—15:00	23:00—24:00
5:00—6:00		15:00—16:00	
6:00—7:00		16:00—17:00	

降雨在工作日各小时的分布情况如表 V3 所示，可见一天的不同时段都同时包含降雨和未降雨的记录。因此，本文的样本可以支持的异质性分析，讨论每个需求情况接近的组内，降雨触发的动态调价对驾驶员劳动参与的影响，并进行组间比较。

表 V3 工作日个小时降雨分布

小时	不下雨的天数	下雨的天数	总天数
0:00—1:00	4	1	5
1:00—2:00	3	2	5
2:00—3:00	3	2	5
3:00—4:00	4	1	5
4:00—5:00	3	2	5
5:00—6:00	3	2	5
6:00—7:00	4	1	5
7:00—8:00	3	2	5
8:00—9:00	3	2	5
9:00—10:00	3	2	5
10:00—11:00	3	2	5
11:00—12:00	3	2	5
12:00—13:00	4	1	5
13:00—14:00	3	2	5
14:00—15:00	5	0	5
15:00—16:00	5	0	5
16:00—17:00	3	2	5
17:00—18:00	3	2	5
18:00—19:00	3	2	5
19:00—20:00	3	2	5
20:00—21:00	2	3	5
21:00—22:00	3	2	5
22:00—23:00	4	1	5
23:00—24:00	4	1	5
总计	81	39	120

附录 VI 临时休息决策

值得注意的是小时劳动供给模型的被解释变量 $D_{idh} = 0$ 可以对应驾驶员临时退出劳动力市场、进行短暂休息，或结束班次，班次劳动供给模型的被解释变量 $Stop_{idt}(\kappa) = 1$ 表示后者。定义驾驶员 i 在日期 d 时间 t 完成行程 κ 后是否临时退出劳动力市场的虚拟变量 $Break_{idt}(\kappa)$ ，行程之间的间隔大于 1 且小于班次划分标准则取值为 1，小于 1 则取值为 0，将模型二中的被解释变量由 $Stop_{idt}(\kappa)$ 替换为 $Break_{idt}(\kappa)$ ，检验降雨对临时暂停劳动供给的影响。

将模型二的被解释变量从停止工作的决策更换为短暂休息的决策后，其估计结果见表 VII。降雨对网约车临时停止工作的意愿影响为负向，对巡游车临时停止工作的意愿不明显，下雨天网约车驾驶员相对于巡游车驾驶员的劳动供给意愿显著提升。

表 VII 临时休息决策的实证结果

	网约车			巡游车		
	<i>break</i>	<i>break</i>	<i>break</i>	<i>break</i>	<i>break</i>	<i>break</i>
	(1)	(2)	(3)	(4)	(5)	(6)
<i>R</i>	-0.003**	-0.006***	-0.006***	0.002	0.002	0.002
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
<i>Y</i> <i>h</i>	√			√		
$\sum \lambda_j I[Y(\kappa) \in \hat{Y}_j]$ $\sum \sigma_i I[h(\kappa) \in \hat{H}_i]$		√			√	
$\sum \lambda_{j,l} I[Y(\kappa) \in \hat{Y}_j] I[h(\kappa) \in \hat{H}_l]$			√			√
控制变量	√	√	√	√	√	√
个体固定效应	√	√	√	√	√	√
日期固定效应	√	√	√	√	√	√
<i>N</i>	225,747	225,747	225,747	197,033	197,033	197,033
<i>R</i> ²	0.102	0.104	0.106	0.109	0.109	0.110

注：***、**、* 分别表示在 1%、5% 和 10% 的水平上显著。被解释变量为是否临时退出劳动力市场的虚拟变量。其中，第（1）—（3）为网约车驾驶员模型二的实证结果，第（4）—（6）列为巡游车驾驶员模型二的实证结果。第（1）、（4）列对应式（2），第（2）、（5）列对应式（3），第（3）、（6）列对应式（4）的回归结果。累积收入的单位为百元，累积时长的单位为小时，式（3）和（4）中累积收入以每 150 元为区间、累积工作时长以每 2 小时为区间分别设置 11 个区间的虚拟变量。括号内均为聚类到班次层面的稳健标准误。

附录 VII 稳健性分析一：区分全职和兼职驾驶员

网约车驾驶员可主要分为全职和兼职两种类型。兼职驾驶员的劳动供给模式可能会不同于全职，并干扰估计结果，为此需要排除兼职驾驶员的影响。鉴于网约车平台没有对全职和兼职驾驶员的清晰界定，本文通过机器学习无监督分类方法对驾驶员进行聚类分析。根据一周内每小时的载客时长，可以得到驾驶员在 168 个小时的行为特征，基于这 168 维信息，本文采用 K 均值聚类方法对网约车驾驶员进行类型划分。

利用 K 均值聚类方法，选择将网约车驾驶员划分为 4 种类型后，不同类型的网约车驾驶员在一周各小时工作人数占比的变化如图 VIII1 所示。从图 VIII1 可以看出，第三组驾驶员在一天中的大部分时间均活跃在出租车市场，因此被界定为全职驾驶员。第二组和第四组驾驶员分别主要活跃于早高峰和晚高峰时段。第一组驾驶员在一周内各小时工作的比例均不高，可被视为兼职驾驶员。

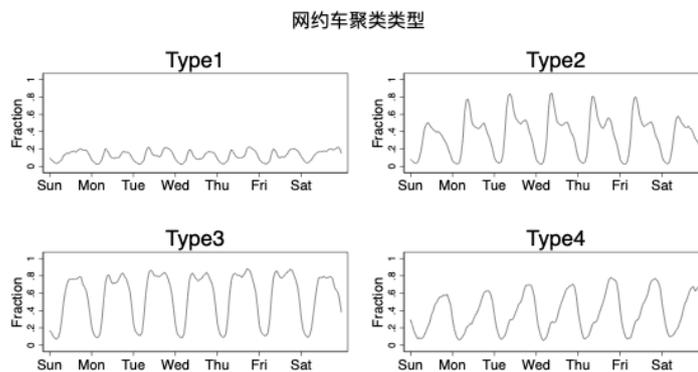


图 VIII1 网约车的聚类类型

将第三组驾驶员界定为全职网约车驾驶员之后，将筛选出来的全职网约车样本与巡游车样本混合，利用二者在非降雨时段的行为特征，同样通过 K 均值聚类方法，筛选出与全职网约车驾驶员最大程度被匹配在同一组的巡游车驾驶员。聚类后网约车和巡游车的分组如表 VIII1 所示。

表 VIII1 全职网约车与巡游车聚类类型

类型	类型 1	类型 2	类型 3	类型 4
Panel A				
巡游车	1420	1216	1411	183
网约车	15	8	38	830
Panel B				
巡游车	173	1484	1349	1224
网约车	828	39	19	5

表 VIII1 中，Panel A 为去掉周一、二、四这些降雨主要出现的日期，仅用周三、五、六、七的小时工作信息对全职网约车和巡游车进行聚类的分组结果；Panel B 为去掉周一、二、四和工作日高峰期的样本，用周三、五非高峰期和周六、日的小时工作信息对全职网约车和巡游车进行聚类的分组结果。从表 VIII1 可以看出，Panel A 中，全职网约车驾驶员主要被分在第四组；Panel B 中，全职网约车驾驶员主要被分在第一组；此外，可以看出，两种分组

方式下全职驾驶员的匹配数量较为相似，体现出匹配结果的稳健性。因此分别利用 Panel A 中的第四组样本和 Panel B 中的第一组样本进行实证模型一的回归，研究降雨对全职驾驶员劳动供给的影响，以排除不同类型驾驶员对识别的干扰。

降雨对全职驾驶员劳动供给的影响的回归结果见表 VII2，其中，第（1）、（2）列为利用表 VII1 的 Panel A 中所匹配全职驾驶员样本，第（3）、（4）列为利用表 VII1 的 Panel B 中所匹配全职驾驶员样本估计的结果。第（1）、（3）列为线性概率模型，第（2）、（4）列为 Logit 模型。从表 VII2 可以看出，在排除了下雨天动态调价可能吸引兼职驾驶员工作对识别带来的干扰后，回归结果显示，在不同的匹配方式下， R 与 $Online$ 交互项前的系数均为正，下雨天全职网约车驾驶员相较同类巡游车驾驶员的劳动供给显著提升。

表 VII2 降雨对全职驾驶员劳动供给的影响

	D	$logit_D$	D	$logit_D$
	(1)	(2)	(3)	(4)
R	-0.029*** (0.008)	-0.169*** (0.036)	-0.023*** (0.008)	-0.125*** (0.039)
$R \cdot Online$	0.053*** (0.009)	0.357*** (0.039)	0.042*** (0.009)	0.267*** (0.041)
控制变量	√	√	√	√
个体固定效应	√	√	√	√
日期固定效应	√	√	√	√
N	105,231	105,231	118,551	118,551
R^2	0.416	0.346	0.376	0.298

注：***、**、* 分别表示在 1%、5% 和 10% 的水平上显著。其中，第（1）、（2）列与第（3）、（4）列全职网约车和巡游车的匹配方式不同，第（1）、（3）列为线性概率模型，第（2）、（4）列为 Logit 模型。除第（2）、（4）列外，括号内均为聚类到驾驶员层面的稳健标准误， R^2 汇报数值在（2）、（4）列为 Pseudo R^2 。

附录 VIII 稳健性分析二：高峰时段动态调价的影响

除了下雨，工作日的早、晚高峰时段也是需求激增、触发动态调价的典型时段。将工作日（周一到周五）的 7:00—10:00 和 17:00—19:00 分别界定为早、晚高峰，式（5）的实证结果如表 VIII1 所示。其中，第（1）—（4）列被解释变量为劳动供给，第（1）、（2）列为线性概率模型，第（3）、（4）列为 Logit 模型，第（5）、（6）列被解释变量为工资；偶数列相较于奇数列增加了个体固定效应和日期固定效应。实证结果显示，交互项前的系数均为正，即降雨和早晚高峰所触发的动态调价均会提升网约车的工资水平和劳动供给。

表 VIII11 降雨和高峰时段对劳动供给的影响

	<i>D</i>	<i>D</i>	<i>logit_D</i>	<i>logit_D</i>	ω	ω
	(1)	(2)	(3)	(4)	(5)	(6)
<i>R</i>	-0.008*	-0.003	-0.043***	-0.011	-0.012***	-0.015***
	(0.005)	(0.002)	(0.007)	(0.009)	(0.004)	(0.005)
<i>Online</i>	0.210***		1.013***		-0.285***	
	(0.004)		(0.009)		(0.016)	
<i>R·Online</i>	0.027***	0.027***	0.152***	0.155***	0.014	0.023**
	(0.006)	(0.006)	(0.015)	(0.017)	(0.012)	(0.011)
<i>Peak·Online</i>	0.114***	0.114***	0.378***	0.391***	0.190***	0.167***
	(0.006)	(0.006)	(0.020)	(0.021)	(0.012)	(0.012)
控制变量	√	√	√	√	√	√
个体固定效应		√		√		√
日期固定效应		√		√		√
<i>N</i>	612,666	612,666	612,666	607,626	234,364	234,352
<i>R</i> ²	0.139	0.199	0.119	0.111	0.046	0.182

注：***、**、* 分别表示在 1%、5% 和 10% 的水平上显著。其中，第（1）—（4）列被解释变量为劳动供给，第（1）、（2）列为线性概率模型，第（3）、（4）列为 Logit 模型，第（5）、（6）列被解释变量为工资；偶数列相较于奇数列增加了个体固定效应和日期固定效应。除第（4）列外，括号内均为聚类到驾驶员层面的稳健标准误，*R*² 汇报数值在（3）、（4）列为 Pseudo *R*²。

附录 IX 稳健性分析三：不同时段而异质性分析

动态调价对驾驶员劳动参与的影响，可能受到其他与小时相关的因素的干扰。为了排除这一干扰因素，本文进一步将工作日的一天 24 个小时分为四个打车需求比较接近的时段。在每组内研究降雨对驾驶员劳动参与的影响，实证结果如表 IX1 所示。其中，第（1）、（2）列表示 0:00—7:00 的凌晨时段，第（3）、（4）列表示 7:00—10:00 和 17:00—19:00 的早、晚高峰时段，第（5）、（6）列表示 10:00—17:00 的日间平峰时段，第（7）、（8）列表示 19:00—24:00 的夜间时段；奇数列列为线性概率模型，偶数列列为 Logit 模型。

R 与 $Online$ 交互项前的系数体现降雨所触发的动态调价在不同时段内对劳动参与的激励作用，在各个时段的影响均为正但略有差异，主要表现为在 0:00—7:00 的凌晨时段以及早晚高峰时段的影响较小。这或许是因为，闲暇在不同时段的“价格”不同，凌晨时段的调价并不足以激励网约车驾驶员增加劳动供给；另一方面是因为高峰时段网约车供给已达极限，难以进一步提升。

表 IX1 不同时段调价对驾驶员影响的实证结果

	0:00—7:00		7:00—10:00 & 17:00—19:00		10:00—17:00		19:00—24:00	
	D	$logit_D$	D	$logit_D$	D	$logit_D$	D	$logit_D$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R	-0.011*** (0.003)	-0.062** (0.026)	-0.012** (0.005)	-0.057*** (0.018)	-0.011** (0.005)	-0.058*** (0.016)	0.001 (0.003)	0.007 (0.023)
$R \cdot Online$	0.014*** (0.003)	0.079*** (0.027)	0.016** (0.008)	0.087*** (0.020)	0.032*** (0.008)	0.208*** (0.019)	0.024*** (0.004)	0.159*** (0.021)
控制变量	√	√	√	√	√	√	√	√
个体固定效应	√	√	√	√	√	√	√	√
日期固定效应	√	√	√	√	√	√	√	√
N	322,811	235,906	230,575	212,476	322,811	286,656	230,581	198,306
R^2	0.221	0.094	0.263	0.008	0.299	0.003	0.312	0.015

注：***、**、* 分别表示在 1%、5% 和 10% 的水平上显著。其中，奇数列列为线性概率模型，偶数列列为 Logit 模型。奇数列括号内均为聚类到驾驶员层面的稳健标准误， R^2 汇报数值在偶数列列为 Pseudo R^2 。

注：该附录是期刊所发表论文的组成部分，同样视为作者公开发表的内容。如研究中使用该附录中的内容，请务必在研究成果上注明附录下载出处。