Electronic Companion for "AI On Time? Evidence from Curb-to-Gate Facial Recognition"

Xiang Wan, wan.207@osu.edu

Hongshuang (Alice) Li, li.815@osu.edu

Zenan Zhou, zzhou119@asu.edu

Web Appendix A: Propensity Score Matching (PSM)

To mitigate potential selection bias, we apply PSM to pair routes in the treatment group with those in the control group. The statistics for variables of the matched routes *before* the implementation of facial recognition are summarized in Table A1 and differ from the statistics for the entire data window reported in Table 1. The latter includes both pre- and post-implementation periods. The routes in the matched treatment group are similar to those in the matched control group regarding all matching variables. The t-tests for the difference between routes in the two groups indicate no statistically significant difference.

Table A1 Matched Routes by PSM

	Matched Treated Routes		**Matched Control Routes**		**Difference (t-test)**	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**t-statistic**	**p-value**
Number of Routes	43		43			
Haul	2691.653	1858.229	2686.768	1124.214	0.014	0.988
Number of airlines	3.209	2.065	2.837	1.572	0.940	0.349
Age of airplanes	15.448	5.363	16.235	5.042	-0.701	0.485
Daily seats	300.916	169.873	315.378	161.021	-0.405	0.686
Daily flights	1.316	0.959	1.307	0.894	0.048	0.962

Web Appendix B: Time Trend and Placebo Tests

We rule out the possibility that the changes in late departures resulted from some consistent time trends between the treatment and control groups before the implementation of facial recognition. To test this, we use the data before facial recognition to estimate the following model:

$$OnTime_{ijt} = \alpha_0^t + \alpha_1^t Treat_{ij} \times Trend_t + Z_{ijt}A^t + Origin_j + Dest_j \qquad \text{(B1)}$$

$$+Year_t + Week_t + Day_t + Hour_t + \varepsilon_{ijt} \qquad ,$$

where $Treat_{ij}$ indicates the flights in the treatment group and $Trend_t$ is an index variable that takes ascending values of 1, 2, 3, . . . for days before facial recognition.

As shown in Table B1, these estimated coefficients are not significant at the 0.05 level. This suggests parallel time trends in late departure between the treatment and control groups before the implementation of facial recognition. Thus, the changes in late departures did not result from pre-implementation time trends.

Table B1 Tests for Time Trend Before the Implementation of Facial Recognition

	On-Time Departures
Variables	$LateDep_{ijt}$
$Treat_{ij} \times Trend_t$	0.001
	(0.001)

Note: *p < 0.05, **p < 0.01, ***p < 0.001.
The fixed effects and all control variables are included in the estimations.

We also conduct placebo tests to assess whether our findings could be replicated by assigning facial recognition to dates other than the actual implementation date. Placebo dates are assigned to every day before and after Nov. 29, 2018. We use the data pre- and post- the actual facial recognition implementation to estimate Model (1) with the placebo dates by replacing $Face_{ijt}$ with $PlaceboFace_{ijt}$. We then estimated the treatment effects using these placebo treatment and control groups. Since the estimated results for the placebo dates include hundreds of coefficients (one for each placebo date), they are not reported in detail. Among all these placebo tests, 96% of the estimated coefficients for placebo dates are not significant at the 0.001 level, and the remaining 4% are significant but have smaller magnitudes than the treatment effect of the actual implementation date. Furthermore, we conduct an alternative placebo test by randomly assigning placebo treated routes and placebo control routes in the control group. Specifically, we randomly select 20 routes as the placebo treated group, and another 20 as the placebo control group. We then re-estimate Model (1) using these placebo treatment and control groups. These estimated coefficients are not statistically significant at the 0.05 significance level. Therefore, the randomly selected placebo treated and control routes do not replicate the significant treatment effect.

Web Appendix C: Changes in Taxi Time and On-Time Arrival

We calculate on-time arrivals similarly to on-time departures, such as

$$LateArr_{ijt} = \text{Max}\left[AcutalArrival_{ijt} - ScheduledArrival_{ijt}, 0\right]$$

$$EarlyArr_{ijt} = \text{Max}\left[ScheduledArrival_{ijt} - AcutalArrival_{ijt}, 0\right] .$$

We replace the dependent variable in Model (1) with $LateArr_{ijt}$, $EarlyArr_{ijt}$, $TaxiOut_{ijt}$, and $TaxiIn_{ijt}$. The descriptive statistics of these variables are reported in Table 1. Similar to on-time departures, due to the skewed distribution, models for $LateArr_{ijt}$ and $EarlyArr_{ijt}$ are estimated using Negative Binomial regressions. The models for $TaxiOut_{ijt}$ and $TaxiIn_{ijt}$ are estimated using Least Squares regression with dummy variables (LSDV) to capture fixed effects. Results are reported in Table 5.

Web Appendix D: Synthetic Control Method

In addition to propensity score matching, we apply the synthetic control method (Abadie et al., 2010, 2015) for the treatment effects of facial recognition. To generate synthetic routes, we aggregate observations to the route-week level. There are 72 routes in the treatment group and 59 in the control group. We construct a synthetic control route for each treated route by calculating the weighted average of routes in the control group. The following variables are used to construct the synthetic routes: the great circle distance between origin and destination, the number of operating airlines, the average age of airplanes, and the average weekly seats available. The descriptive statistics for treated and synthetic routes are reported in Tables D1 and D2, where the values in the two groups are close, and t-tests suggest no significant difference.

Table D1 Characteristics of Treatment Routes and Synthetic Routes for Late Departure

	Late Departure			
	Treated Routes	**Synthetic Route**	**Difference**	
Variables	**Mean**	**Mean**	**t-statistic**	**p-value**
Weekly Average Haul	2,242	2,195	0.147	0.884
Weekly Average Number of Airlines	3.075	3.139	-0.181	0.857
Weekly Average Age of Airplanes	15.878	15.433	0.448	0.655
Weekly Total Seats	2,288	2,253	1.021	0.310

Table D2 Characteristics of Treatment Routes and Synthetic Routes for Early Departure

	Early Departure			
	Treated Routes	**Synthetic Route**	**Difference**	
Variables	**Mean**	**Mean**	**t-statistic**	**p-value**
Weekly Average Haul	2,242	2,247	-0.002	0.999
Weekly Average Number of Airlines	3.075	3.165	-0.263	0.793
Weekly Average Age of Airplanes	15.878	14.918	0.868	0.388
Weekly Total Seats	2,288	2,266	0.642	0.522

We follow the DID framework to estimate the treatment effects of facial recognition. The estimated results are reported in Table D3, which are consistent with results using PSM. That is, facial recognition technology reduces late departure but does not significantly impact early departure.

Table D3 Treatment Effects on On-Time Departure Using the Synthetic Control Method

Variables	Late Departure	Early Departure
$Face_{jt}$	-0.119***	0.065
	(0.032)	(0.040)

*** p<0.001, ** p<0.01, * p<0.05.
Fixed effects and control variables are included in model estimations.

Web Appendix E: Event Study Method

In the event study method, we first use the flight data in the matched treatment group before the implementation of facial recognition to estimate the following model for late and early departure. The treatment variable $Face_{ijt}$ is not included because the data in the control group are not included in the event study. The estimated results are reported in Table E1.

$$OnTime_{ijt} = \alpha_0 + \alpha_1 Seat_{ijt} + \alpha_2 Haul_{ijt} + \alpha_3 PlaneAge_{ijt} + \alpha_4 Precipitation_{ijt} \quad \text{(E1)}$$
$$+\alpha_5 Snow_{ijt} + \alpha_6 MaxTemp_{ijt} + \alpha_7 MinTemp_{ijt} + \alpha_8 Wind_{ijt} + \alpha_9 Connect_{jt}$$
$$+\alpha_{10} Congestion_{jt} + \alpha_{11} AsiaAfrica_j + Year_t + Week_t + Day_t + Hour_t + \varepsilon_{ijt}$$

Table E1. The Estimated Results of the On-Time Departure before Facial Recognition

Variables	$LateDep_{ijt}$	$EarlyDep_{ijt}$
$Seat_{ijt}$	0.002**	-0.002***
	(0.001)	(0.000)
$Haul_j$	0.000	0.001*
	(0.000)	(0.000)
$PlaneAge_{ijt}$	0.008*	-0.003**
	(0.004)	(0.001)
$Precipitation_{ijt}$	0.017***	-0.010***
	(0.002)	(0.001)
$Snow_{ijt}$	0.028**	-0.007*
	(0.009)	(0.003)
$MaxTemp_{ijt}$	-0.011	0.019**
	(0.010)	(0.006)
$MinTemp_{ijt}$	-0.007	-0.008
	(0.010)	(0.006)
$Wind_{ijt}$	0.067**	-0.027**
	(0.023)	(0.013)
$Connect_{jt}$	0.001***	-0.000
	(0.000)	(0.000)
$Congestion_{jt}$	0.019***	-0.021**
	(0.004)	(0.008)
$AsiaAfrica_j$	0.020***	0.076
	(0.003)	(0.051)
Constant	1.948***	2.760***
	(0.213)	(0.038)
Fixed Effects for Year	included	included
Fixed Effects for Week	included	included
Fixed Effects for Day	included	included
Fixed Effects for Hour	included	included
Observations	15,881	15,881
Log-Likelihood	-40,300	-26,837

*** p<0.001, ** p<0.01, * p<0.05

Then, we use the estimated coefficients in Table E1 to predict the on-time departure after the facial recognition implementation. The abnormal values are calculated by the difference between the observed on-time departure and the predicted on-time departure after facial recognition. The mean value of abnormal on-time departure is calculated as follows:

$$\mu_{abnormal\ on\_time} = \frac{\sum(Observed\ OnTime_{ijt} - Predicted\ OnTime_{ijt})}{N}$$

As shown in Table E2, the average abnormal value for late departure is -1.497, and the t-test suggests it is significantly different from zero. It indicates that facial recognition technology helps reduce late departures, resulting in a 10.671% improvement (1.497/14.029×100%=10.671%). In contrast, the average abnormal value for early departure is 0.019, and the t-test shows it is not significantly different from zero.

Table E2. Abnormal On-Time Departure in the Event Study

	Abnormal Mean	**Observed Mean**	**Predicted Mean**	**Confidence Interval (95%)**	**T-tests**
$LateDep_{ijt}$	-1.497	12.326	14.029	(-2.082, -0.911)	-5.009***
$EarlyDep_{ijt}$	0.019	1.901	1.868	(-0.019, 0.056)	0.985

*** p<0.001, ** p<0.01, * p<0.05

While the event study method analyzes observations in the treated group pre- and post-implementation of facial recognition, DID captures both the differences between the pre- and post-implementation periods as well as the differences between the treatment and control groups. Thus, we report the DID analyses in the main text.

Web Appendix F: Treatment Effect in the Sample with No Matching, 1-to-m Matching, and Alternative Estimation Methods

To ensure our results remain robust without applying PSM, we estimate models using the data of all flights before PSM. Meanwhile, in addition to 1-to-1 matching reported in the main results, we estimate models using data after 1-to-m matching. The results are reported in Table F1. They are consistent with the main results in the paper: departure delay decreased after facial recognition was implemented. To ensure that our results are not driven by the functional forms employed (negative binomial), we estimate models again with Tobit and Poisson estimation. The estimated coefficients for $Face_{ijt}$ are also reported in Table F1. The signs and significance levels of these coefficients are consistent with our main results.

Table F1 Treatment Effects Using Data with No Matching, 1-to-m Matching, and Alternative Estimation for On-Time Performance

	Flight On-Time Records							
	Data with No Matching		Data with 1-to-m Matching		Tobit Regression		Poisson Regression	
Variables	$LateDep_{ijt}$	$EarlyDep_{ijt}$	$LateDep_{ijt}$	$EarlyDep_{ijt}$	$LateDep_{ijt}$	$EarlyDep_{ijt}$	$LateDep_{ijt}$	$EarlyDep_{ijt}$
$Face_{ijt}$	-0.112***	0.023	-0.134***	0.039	-1.410***	0.065	-0.122***	0.017
	(0.008)	(0.020)	(0.011)	(0.023)	(0.341)	(0.042)	(0.016)	(0.013)

*** p<0.001, ** p<0.01, * p<0.05. Fixed effects and control variables are included in all model estimations

Web Appendix G: Changes in Seat Capacity, Flight Frequency, Turnaround, and Block Time

Although we find a significant improvement in late departures after facial recognition, changes in on-time departure could be caused by alternative mechanisms, such as the number of seats, flight frequency, turnaround time, or block time.

First, passengers may avoid the technology by switching to other terminals or other airlines. As a result, the number of passengers departing from the terminal with facial recognition may largely decrease. If so, the improved on-time performance may result from reduced passenger numbers. If the passenger number significantly declined after the implementation of facial recognition, the airlines would arrange airplanes with smaller seat capacity for these flights. Second, airlines may schedule fewer flights departing from the terminal equipped with facial recognition to magnify the improved boarding efficiency and justify the implementation decision to adopt facial recognition (or schedule fewer flights to minimize the potential risk of trying out the new technology). With fewer flights scheduled to depart from a terminal, the on-time performance could be affected by the availability of shared terminal resources (such as gates, crews at gates, air traffic controllers, flight dispatchers, airplane mechanics, airplane fuelers, and workers). Thus, the changes in flight frequency from an airport terminal could also be a reason for the changes in on-time departure.

To rule out these alternative mechanisms, we examine the changes in seat numbers (as an indicator of passenger numbers) and flight frequency after facial recognition. Specifically, we aggregate the data to the terminal-date level and then estimate the DID models with the daily seat numbers and the daily number of flights as the dependent variables.

$$Alternative_{at} = \alpha_0 + \alpha_1 Face_{at} + Terminal_a + Year_t + Week_t + Day_t + \varepsilon_{at} \qquad \text{(G1)}$$

where $Alternative_{at}$ is the dependent variable for terminal a on day t, representing the daily seat number ($Seat_{at}$) and the daily flight frequency ($FlightFreq_{at}$). The independent variable is $Face_{at}$, which is coded 1 for the terminal a on day t when the facial recognition is available and 0 otherwise. Fixed effects are included to control for unobservable variations at the departure terminal ($Terminal_a$), such as their infrastructures, congestion, operations, and teams. Fixed effects for departure year ($Year_t$), week of the year ($Week_t$), and day of the week (Day_t) are included.

We estimate the models using LSDV. The natural logarithmic transformation is applied to the dependent variables for a closer approximation of a Normal distribution. The estimated results are reported in Table G1. None of the estimated coefficients are significant. Our results suggest that seat capacity and flight frequency did not change significantly after implementing facial recognition. Therefore, changes in on-time departure do not result from these potential alternative mechanisms.

Table G1 Changes in Flight Seat Capacity and Flight Frequency

Dependent Variable	$\ln Seat_{at}$	$\ln FlightFreq_{at}$
$Face_{at}$	0.006	-0.023
	(0.009)	(0.014)
Constant	9.248***	3.853***
	(0.036)	(0.036)
Fixed Effects for Terminal	included	included
Fixed Effects for Year	included	included
Fixed Effects for Week	included	included
Fixed Effects for Day	included	included
R-Squared	0.806	0.768
Adjusted R-Squared	0.798	0.758

Note: *** p<0.001, ** p<0.01, * p<0.05

In addition, airlines could potentially improve on-time departures by increasing aircraft turnaround time or increase on-time arrivals by increasing flight block time. To rule out these alternative explanations, we also analyze the changes in turnaround and block time after facial recognition. The estimated results are reported in Table G2. The treatment effects on turnaround and block time are not statistically significant, indicating that they did not change significantly after implementing facial recognition. Therefore, changes in on-time departure and arrival did not result from changes in turnaround and block time.

Table G2 Changes in Turnaround and Block Time

Dependent Variable	ln$TurnAround_{ijt}$	ln$BlockTime_{ijt}$
$Face_{ijt}$	-0.029	0.001
	(0.019)	(0.001)
Constant	1.170***	5.149***
	(0.221)	(0.196)
Fixed Effects for Airport	included	included
Fixed Effects for Year	included	included
Fixed Effects for Week	included	included
Fixed Effects for Day	included	included
R-Squared	0.438	0.689
Adjusted R-Squared	0.431	0.686

Note: *** p<0.001, ** p<0.01, * p<0.05

Web Appendix H: Countries With No Biometric Passports

During the time window of the study, the countries in our data that did not use biometric passports include Guatemala, South Africa, Nicaragua, Dominican Republic, El Salvador, Honduras, Ghana, India, and Cuba. To rule out this potential explanation of biometric passports, we estimate the models using data after excluding these countries that do not use biometric passports. The estimated results are reported in Table H1. The results are consistent with the main findings, indicating that the observed difference in the impact of facial recognition on on-time performance between flights to Asia and Africa and those to other regions cannot be attributed to variations in biometric passport usage.

Table H1 Treatment Effect on Late Departures
Excluding Destination Countries with No Biometric Passports

	Model (1)	Model (2)
Variables	$LateDep_{ijt}$	$LateDep_{ijt}$
$Face_{ijt}$	-0.135***	-0.109***
	(0.009)	(0.006)
$AsiaAfrica_j$	0.066***	0.062***
	(0.013)	(0.017)
$Face_{ijt} \times AsiaAfrica_j$		0.043***
		(0.013)
Constant	2.840***	2.846***
	(0.514)	(0.531)
Fixed Effects for Airports	included	included
Fixed Effects for Year	included	included
Fixed Effects for Week	included	included
Fixed Effects for Day	included	included
Fixed Effects for Hour	included	included
Observations	56,918	56,918
Log-Likelihood	-127,592	-127,590

*** p<0.001, ** p<0.01, * p<0.05