SUPPLEMENT TO "MARKET STRUCTURE AND MULTIPLE EQUILIBRIA IN AIRLINE MARKETS"

(*Econometrica*, Vol. 77, No. 6, November 2009, 1791–1828)

BY F. CILIBERTO AND E. TAMER

We first describe the simulation procedure. Then we give theorems that are used in the main text. We also include many of the practical details that are used to implement the estimation procedure.

KEYWORDS: Entry models, inference in discrete games, multiple equilibria, partial identification, airline industry, firm heterogenelty.

S1. SIMULATION PROCEDURE

WE SIMULATE the functions $\mathbf{H}_1(\mathbf{X}, \theta)$ and $\mathbf{H}_2(\mathbf{X}, \theta)$ for a given $\mathbf{X}$ and θ as follows. Set $\widehat{\mathbf{H}}_1(\mathbf{X}, \theta) = \widehat{\mathbf{H}}_2(\mathbf{X}, \theta) = 0$. Moreover, for every market, generate and store R draws from the distribution F with identity variance–covariance matrix. The number of simulations is assumed to go to infinity with sample size. More on this below. For each simulation $r = (1, \dots, R)$, follow Steps 1–3:

STEP 1: Transform the given matrix of epsilon draws into a draw with covariance matrix specified in θ. This is stored in $\boldsymbol{\epsilon}^r$.[1]

STEP 2: Using the profit function from (1), calculate

$$\boldsymbol{\pi}(\mathbf{y}_j, \mathbf{X}, \theta, \boldsymbol{\epsilon}^r) = [\pi_1(\mathbf{y}_{-1}, \mathbf{X}, \theta, \epsilon_1^r), \dots, \pi_K(\mathbf{y}_{-K}, \mathbf{X}, \theta, \epsilon_k^r)]$$

for all $j = 1, \dots, 2^K$.

STEP 3: Find the equilibria of the game:

- For all $j \in \{1, \dots, 2^K\}$ such that $\boldsymbol{\pi}(\mathbf{y}_j, \mathbf{X}, \theta, \boldsymbol{\epsilon}^r) \geq 0$, set $\widehat{H}_2^j = \widehat{H}_2^j + 1$.
- If there is a $j \in \{1, \dots, 2^K\}$ such that $\boldsymbol{\pi}(\mathbf{y}_j, \mathbf{X}, \theta, \boldsymbol{\epsilon}^r) \geq 0$ *uniquely*, that is, there is no $j' \neq j$ such that $\boldsymbol{\pi}(\mathbf{y}_{j'}, \mathbf{X}, \theta, \boldsymbol{\epsilon}^r) \geq 0$, then $\widehat{H}_1^j = \widehat{H}_1^j + 1$.

This will provide us with the simulated versions

$$\frac{1}{R}\widehat{\mathbf{H}}_2(\mathbf{X}, \theta) \quad \text{and} \quad \frac{1}{R}\widehat{\mathbf{H}}_1(\mathbf{X}, \theta).$$

[1]There are many ways to do this, one of which is to obtain the Cholesky decomposition of the given covariance matrix and use it to transform independent draws into dependent draws.

 DOI: 10.3982/ECTA5368

S2. CONSISTENCY, PRACTICAL ESTIMATION, AND CONFIDENCE REGIONS

In this section, we describe procedures to construct regions that cover the identified set with a prespecified probability. We also describe confidence regions for the identified parameter.

First Stage Estimation of Choice Probabilities: Our minimum distance estimator calls for estimating the choice probability vector $\mathbf{P}(\mathbf{x}) = \mathbf{P}(\mathbf{y}|\mathbf{X} = \mathbf{x})$ used in (6) in a first step. We can use a nonparametric conditional expectation estimator to obtain this estimator, $\mathbf{P}_n(\mathbf{x})$. The CHT theory that is developed to obtain confidence regions for sets relies on having a *finite number* of moment inequalities; hence we assume that the data have finitely many support points (discrete support) or that

$$X \in S_x = \{x_1, \dots, x_d\}. \tag{S1}$$

We use a simple frequency estimator to get the conditional choice probabilities:

$$P_n^{(y')}(x) = \frac{\sum_i 1[y_i = y']1[x_i = x]}{\sum_i 1[x_i = x]}.$$

It is easy to see that in this case

$$\begin{aligned}
&\sup_x \left|P_n^{(y')}(x) - P^{(y')}(x)\right| \\
&\quad = \max_x \left\{P_n^{(y')}(x_1) - P^{(y')}(x_1), \dots, P_n^{(y')}(x_d) - P^{(y')}(x_d)\right\} \\
&\quad = o_p(1).
\end{aligned}$$

The objective function we use again is

$$\begin{aligned}
Q(\boldsymbol{\theta}) &= \int \left[\left\|(P(x) - H_1(x, \theta))_-\right\| + \left\|(P(x) - H_2(x, \theta))_+\right\|\right] dF_x \\
&= \sum_{j=1}^{d} p_j \left[\left\|(P(x_j) - H_1(x_j, \theta))_-\right\| + \left\|(P(x_j) - H_2(x_j, \theta))_+\right\|\right],
\end{aligned}$$

where $(A)_- = [a_1 1[a_1 \le 0], \dots, a_{2^k} 1[a_{2^k} \le 0]]$ and similarly for $(A)_+$ for a 2^k vector A, $\|\cdot\|$ is the Euclidian norm, and p_j is the probability conditional on $X = x_j$. It is easy to see that $Q(\boldsymbol{\theta}) \ge 0$ for all $\theta \in \Theta$ and that $Q(\boldsymbol{\theta}) = 0$ if and only if $\theta \in \Theta_I$, the identified set in definition (S1) above. The sample analog of the above objective function is

$$Q_n(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{n}\big[\big\|(P_n(x_i) - \widehat{H}_1(x_i,\theta))_-\big\| + \big\|(P_n(x_i) - \widehat{H}_2(x_i,\theta))_+\big\|\big], \tag{S2}$$

where $\|\cdot\|$ is the Euclidian distance and again $P_n(x)$ is the vector of 2^k choice probabilities estimated from the data. The confidence regions are appropriately constructed level sets of the objective function. To apply Theorem 3, we need to check the uniformity results in the statement of the theorem.

Here, we first describe how one obtains a confidence region for the *identified set*. Earlier versions of this paper reported these. Below, we also provide descriptions of how we obtained confidence regions for the (potentially partially identified) true parameter. The theoretical properties of these regions are studied in CHT.

Confidence Region for the Set: A key statistic for building this confidence region is

$$\mathcal{C}_n = \sup_{\theta\in\Theta_I} nQ_n(\theta).$$

This is because our confidence regions are level sets $\mathcal{C}_n(c)$ of the objective function $Q_n(\cdot)$ written as

$$\mathcal{C}_n(c) = \{\theta\in\Theta : nQ_n(\theta)\le c\}. \tag{S3}$$

Hence, a set $\hat{\Theta}_I = \mathcal{C}_n(c)$ covers Θ_I at level α:

$$P(\Theta_I\subseteq\hat{\Theta}_I) = P\Big(\sup_{\theta\in\Theta_I} nQ_n(\theta)\le c\Big) = P(\mathcal{C}_n\le c)$$

if c is chosen as the level-α quantile of $\mathcal{C}_n$. As usual, we use large n asymptotics to approximate this cutoff level. To do that, we need to derive this asymptotic distribution of $\mathcal{C}_n$. First, define the boundary of the set as

$$\begin{aligned}\partial\Theta_I = \big\{\theta_I\in\Theta_I : {}& H_1(x_j;\theta_I) = P(x_j)\text{ or}\\ & H_2(x_j;\theta_I) = P(x_j),\ \text{for some } j\le d\big\}\end{aligned}$$

and let $n_j = \frac{1}{n}\sum_i 1[x_i = x_j]$. Define $\hat{W}_j := \sqrt{n}(P_n(x_j) - P(x_j))$ for $j = 1,\dots,d$. Assume also that the central limit theorem and the law of large numbers apply such that

$$\begin{aligned}&(\hat{W}_1,\dots,\hat{W}_d)\to_d (W_1,\dots,W_d)\sim\mathcal{N}(0,\Omega),\\ &n_j/n\to_p p_j\quad\text{for each } j\le d.\end{aligned} \tag{S4}$$

To deal with the presence of the simulated quantities, we assume that the number of simulations, R, goes to infinity at a rate $R = O(n^{2+\alpha})$, where $\alpha > 0$. This will guarantee that the simulations will not have an effect on the asymptotic distribution and hence can be ignored. Given the above discussion, Assumption 1, and the assumption also that the parameter space is a compact subset of a finite dimensional Euclidian space, one can show, using similar steps as, for example, page 16 of Chernozhukov, Hong, and Tamer (2002) that

$$\mathcal{C}_n \to_d \mathcal{C}, \tag{S5}$$

where

$$\mathcal{C} = \sup_{\theta \in \partial\Theta_I} \sum_{j=1}^{d} (W_j)_+^2 1[P_j = H_1(x_j, \theta)] + (W_j)_-^2 1[P_j = H_2(x_j, \theta)]. \tag{S6}$$

It is hard to simulate the α-quantile of the above statistic since it is not pivotal. We follow CHT and subsample the distribution of $\mathcal{C}_n$ above to obtain an asymptotic approximation to its α-quantile. We use a modified procedure to account for misspecification of the model (where the minimum of the function Q in the population might not be equal or even close to zero). We instead report

$$C_n(c) = \left\{\theta \in \Theta : n\Big(Q_n(\theta) - \min_t Q_n(t)\Big) \le c\right\} \tag{S7}$$

for an appropriately constructed c. First, we construct all subsets B_n of size $b \ll n$. We take b to be equal to $n/4$.[2] We start with an initial value $c_{(0)}$ for the cutoff (see below for a way to choose this $c_{(0)}$). We then compute

$$\widehat{\mathcal{C}}_{i,b,n,c_0} = \sup_{t \in C_n(c_{(0)})} b(Q_b(t) - q_b) = \sup_{t \in C_n(c_{(0)})} b\Big(Q_b(t) - \min_t Q_b(t)\Big) \tag{S8}$$

for each ith subset, $i \le B_n$, where $C_n(c_{(0)}) = \{\theta \in \Theta : n(Q_n(\theta) - \min_t Q_n(\mathbf{t})) \le c_0\}$. This involves minimizing the objective function at each subsample. Here, we use the Nelder–Mead algorithm with a starting value equal to the arg min obtained using the full data set.

Initial Choice of $c_{(0)} = c_0$: The initial choice of the cutoff that we use here is always 25% above the minimum sample objective function value. Starting with this initial choice, we iterate the objective function twice and use that final cutoff level as the quantile that defines our confidence region. We find that iterating further does not change the cutoff by much. We then compute the α-quantile of the numbers $\widehat{\mathcal{C}}_{i,b,n,c_{(2)}}$ which provides appropriate coverage

[2]There is no general theory of picking a subsample size. See Politis, Romano, and Wolf (1999) for more on this point. However, trying different b's in this paper led to similar results.

properties (asymptotically). One can also use $c_{(0)} = 0$ as the starting cutoff. For this class of models, this was shown in CHT to deliver a set with the appropriate confidence property.

Summary of Procedure to Obtain Confidence Region for the Set Θ_I:

S1. We minimize the objective function $Q_n(\cdot)$ using a genetic algorithm that we describe below. In the process, we collect values for the objective function at many ($\approx$ 1,000,000) randomly chosen parameters using this Markov chain Monte Carlo-like procedure.

S2. For every subsample, we minimize the *subsampled* objective function (the one constructed with the subsample as opposed to the full data set) over the initial estimate of the set we constructed in step S1. We then obtain the empirical α-quantile of the set $\{\widehat{\mathcal{C}}_{i,b,n,c_0} : i \leq B_n\}$. That gets us a new cutoff $\hat{c}_{(1)}$. Note that here we need to evaluate q_b at each subsample, which requires an optimization step. We do this using the Nelder–Mead equation starting at the overall minimum found in step S1.

S3. We iterate steps S1 and S2 two times to obtain $\hat{c}_{(2)}$.

S4. We replace c with $\hat{c}_{(2)}$ in (7) to obtain the confidence region we report. It is then easy to prove (similar to Lemma 3.1 of CHT) the following lemma.

LEMMA S2: *Suppose that* (5) *holds where $\mathcal{C}$ is as in* (6). *Then for any $\widehat{c} \to_p c(\alpha) := \inf\{c \geq 0 : P\{\mathcal{C} \leq c\} \geq \alpha\}$ for $\alpha \in (0,1)$, such that $\widehat{c} \geq 0$ with probability* 1, *we have that as $n \to \infty$, $P\{\Theta_I \subseteq C_n(\widehat{c})\} = P\{\mathcal{C}_n \leq \widehat{c}\} = P\{\mathcal{C} \leq c(\alpha)\} + o(1) = \alpha + o(1)$ if $c(\alpha) > 0$ and $P\{\Theta_I \subseteq C_n(\widehat{c})\} = P\{\mathcal{C}_n \leq \widehat{c}\} \geq P\{\mathcal{C} = 0\} + o(1) \geq \alpha + o(1)$ if $c(\alpha) = 0$.*

For more on consistent estimators for the set, see CHT.

Confidence Region for the Point: To obtain confidence regions for the true parameter, we use a slight modification of the above procedure. A confidence region for the point is

$$\widehat{\Theta}_I = \left\{\theta \in \Theta : n\Big(Q_n(\theta) - \min_t Q_n(t)\Big) \leq \min(c_n, c_n(\theta))\right\}, \tag{S9}$$

where $c_n(\theta)$ is a consistent estimate of $c(\theta)$, the α-quantile of $\mathcal{C}(\theta)$ where

$$n\Big(Q_n(\theta) - \min_t Q_n(t)\Big) \to_d \mathcal{C}(\theta).$$

We use subsampling as described on page 1269 of CHT to get $c_n(\theta)$. Here, $c_n(\theta)$ is the α-quantile of $\{b_n(Q_{b_n,j}(\theta_0) - \min_t Q_{b_n,j}(t)), j = 1, \ldots, B_n\}$. We use $\frac{n}{4}$ as the subsample size[3] b_n. It is shown in CHT that the probability that θ_I is

[3]We have experimented with subsample sizes of $\frac{n}{3}$, $\frac{n}{5}$, and $\frac{n}{6}$ with similar results.

in this confidence region is no smaller than

$$P\Big\{n\Big(Q_n(\theta_I) - \min_t Q_n(t)\Big) \le [c(\theta_I) + o_p(1)] \vee 0\Big\}$$
$$= P\{\mathcal{C}(\theta_I) \le c(\theta_I)\} + o(1) \ge \alpha + o(1),$$

which is the desired coverage property. In the text, we report instead

$$C_n(c_2) = \Big\{\theta \in \Theta : n\Big(Q_n(\theta) - \min_t Q_n(t)\Big) \le c_2\Big\},$$

where we have

$$\liminf_{n\to\infty} P\{\theta_I \in C_n(c_2)\} \ge \liminf_{n\to\infty} P\{\theta_I \in \widehat{\Theta}_I\} \ge \alpha$$

for all $\alpha_I \in \Theta_I$. Here, c_n is an estimate of an upper bound on $\sup_{\theta\in\Theta_I} c(\theta)$ as in page 1805 of the main text, and c_2 is the second iteration as described in the text. We report the confidence regions $C_n(c_2)$, since we do not see any practical difference between them and $\widehat{\Theta}_I$ above, and because $C_n(c_2)$, being a level set, is easier to compute. Earlier versions of this paper contained confidence regions on the set, but the current results report $C_n(c_2)$ as the confidence regions, as requested by the co-editor.

Computational Issues: The optimization was done using the canned routine *simulannealbnd* in Matlab. For each specification, we started our search from at least five starting values.[4] This is helpful since genetic algorithms, although slow, scan the surface of the function and thus allow us to obtain the level sets needed to construct our set estimates. From the overall minimum, we run annealing for a while longer (usually a day or two for every specification) to evaluate the functions at many different parameter values close to the minimum we found. This will give us a snapshot of the surface of the function.

One issue that arises when solving for equilibria of a given game is that sometimes the game admits equilibria only in mixed strategies. In principle, it is not conceptually difficult to deal with mixed strategy equilibria since at a particular iteration, one can compute the mixing probabilities. However, as we say in the text, we do not do that here for simplicity, and rather deal with this problem as follows. If a game does not have an equilibrium in pure strategies *for some* realization of the errors in one market, then we do not consider that particular realization of the errors when we construct the lower and upper bounds. If a game does not have an equilibrium in pure strategies for any realization of the errors in one market, then we do not consider that particular market when we construct the lower and upper bounds. In the minimization we keep track of the percentage of realizations and markets where there are no equilibria in

[4]For the simplest specifications, we used more than 20 starting values.

pure strategies. In our data, it *never* occurred that no pure strategy equilibria existed.

In *none* of our searches did we restrict the competitive effects to be positive. This is in sharp contrast to the previous literature, which had to assume a sign on the competitive effects.

We used both 20 and 100 simulations for each market. The results were essentially identical. In both cases, the minimization routine would sometimes diverge to two minima, one of which would not be economically intuitive (e.g., positive competitive effects). We would then increase the number of simulations to 5000 and see that the "unreasonable" minimum would not be a minimum any longer.

To construct the confidence intervals we subsample the data sets with subsample sizes equal to one-fourth of the data. The results did not change much when using subsamples of smaller sizes. We also simulate from the error term in every subsample which guarantees that the simulation error is taken care of.

Moreover, note that subtracting the minimum of the function as in (S8) is essential to guarantee that the confidence regions are nonempty. This is important since we assume throughout that the model is well specified and that the set Θ_I is nonempty.

S2.1. *Data Construction*

Data sets: We use three data sets from the Origin and Destination Survey (DB1B), which is a 10% sample of airline tickets from reporting carriers. The observations are from the first quarter of 1996 to the fourth quarter of 2007. These data are collected by the U.S. Department of Transportation.

The first data set is the DB1B Coupon Origin and Destination Dataset, which provides coupon-specific information for each domestic itinerary of the Origin and Destination Survey, such as the operating carrier, origin and destination airports, number of passengers, fare class, coupon type, trip break indicator, and distance. We merge this data set by operating carrier with the T-100 Domestic Segment Dataset. The T-100 Domestic Segment Dataset contains domestic market data by air carrier, and origin and destination airports for passengers enplaned. The T-100 is not a sample: It reports all flights that occurred in the United States in a given month of the year.

From the merged data set we drop those tickets involving flights that are not provided on a regular basis or for which there is no record in the T-100 segment. We drop all tickets that involve a flight that is not provided at least once a week.

Then we merge by ticket identification numbers the reduced DB1B Coupon Origin and Destination Dataset with the DB1B Market and Ticket Origin and Destination Dataset. The DB1B Market Origin and Destination Dataset contains directional market characteristics of each domestic itinerary of the Origin and Destination Survey, such as the reporting carrier, origin and destination airport, prorated market fare, number of market coupons, market miles flown, and carrier change indicators. The DB1B Ticket contains summary characteristics of each domestic itinerary on the Origin and Destination Survey, including the reporting carrier, itinerary fare, number of passengers, originating airport, roundtrip indicator, and miles flown. The unit of observation in this data set is a ticket.

One important issue is how to treat regional airlines that operate through code-sharing with national airlines. We assume that the decision to serve a spoke is made by the regional carrier, which then signs code-share agreements with the national airlines. As long as the regional airline is independently owned and issues tickets, we treat it separately from the national airline.

Market Definition: We define a market as the trip between two airports, irrespective of intermediate transfer points. Because of data limitations, Berry (1992) defined a market as the market for air passenger travel between two cities, which rules out that demand is different for airports in the same city or metro area. Following Borenstein (1989), we assume that flights to different airports in the same metropolitan area are in separate markets.

Data Cleaning: We drop (i) tickets with more than six coupons; (ii) tickets involving U.S. nonreporting carriers flying within North America (small airlines serving big airlines) and foreign carrier flying between two U.S. points; (iii) tickets that are part of international travel; (iv) tickets involving noncontiguous domestic travel (Hawaii, Alaska, and territories); (v) tickets whose fare credibility is questioned by the Department of Transportation; (vi) tickets that are neither one-way nor round-trip travel; (vii) tickets including travel on more than one airline on a directional trip (known as interline tickets); (viii) tickets with fares less than 20 dollars; (ix) tickets in the top and bottom five percentiles of the year–quarter fare distribution. Finally, Berry (1992) defined a firm as serving a market if it transported at least 90 passengers in one quarter. This corresponds to a once a week flight by a medium size jet. Since we already control for firms that fly less than once a week and since markets can be served by small regional jets, we change the threshold to 20 passengers (each way). We then aggregate the ticket data by ticketing carrier; thus the unit of observation is market–carrier–year–quarter specific.

In this paper we are only interested in knowing whether a carrier served a market. Therefore, the aggregation is straightforward: For each carrier, we construct a categorical variable that is equal to 1 if the carrier serves the market and equal to 0 otherwise. After constructing the categorical variables, the relevant unit of observation is market–year–quarter specific.

Market Selection: To select the markets, we merge this data set with demographic information on population from the U.S. Census Bureau for all the metropolitan statistical areas (MSAs) of the United States. We then construct a ranking of airports by the MSA's market size. The data set includes a sample of markets between the top 50 metropolitan statistical areas, ranked by population size. We exclude the Muskegon County Airport, the Saint Petersburg/Clearwater International Airport, and the Atlantic City International Airport because there are too few markets between these airports and the remaining airports. Including them would increase the number of unserved markets artificially. As mentioned in the text, we include markets that are *temporarily* not served by any carrier. To identify markets that are almost never served by any carrier from markets that are only temporarily not served by any carrier, we proceed as follows. We consider the full 1996–2007 data set of market–carrier–year–quarter observations. For each market, we compute the number of quarters that a market has been served by at least one carrier. Then we drop from the data set those markets that have not been served in at least 50% of all the quarters in the full data set. We keep markets out of and to Dallas Love airport which are at least 500 miles distant from the Dallas airport. This last condition is to investigate the effect of the Wright Amendment on carriers' entry decisions.

Airline Types: We lump some of the carriers in our data set in two types. There are two reasons to do this. First, many low cost carriers are present in only a few markets, and lumping them allows us to use a meaningful grouping that captures the impact of a small low cost carrier presence in the market. Second, the number of possible market structures that can be an equilibrium grows exponentially with the number of firms. For any K firms, there are 2^K possible market structures. This is clearly prohibitive with many firms. We lump Northwest, Continental, America West, and USAir under the medium airline type. To facilitate this assumption, we drop markets where one of the two endpoints is one of these hubs: Minneapolis, Detroit, Memphis, Cleveland, Newark, Houston International, Charlotte, Philadelphia, Pittsburgh, Phoenix, Las Vegas.

Carrier Airport Presence: The construction of the variable *carrier airport presence* is straightforward. For example, when we consider Delta, we proceed as follows. If Delta serves 60 markets out of Atlanta and there are 84 markets that are served out of Atlanta, then for each market that we consider out of Atlanta (e.g., Atlanta to Chicago O'Hare), Delta serves $59/83 \simeq 71\%$ of the other markets out of Atlanta. We repeat the same computation for the other endpoint and then take the average.

The construction of the variable requires some additional steps when we consider types of firms. When we consider the medium airlines (MA), we first compute the airport presence for USAir, Continental, and America West, and then we take the maximum of the three. When we consider the low cost carriers (LCC), we first compute the airport presence of each of the low cost carriers, and then again we take their maximum.

The Opportunity Cost of Serving a Market: To construct the measure of *cost*, we consider the following hub airports: Dallas/Fort Worth and Chicago O'Hare for American; Cleveland, Houston International, and Newark for Continental; Atlanta, Cincinnati, and Dallas/Fort Worth for Delta; Phoenix and Las

Vegas for America West; Minneapolis and Detroit for Northwest; Denver and Chicago O'Hare for United; Charlotte, Pittsburgh, and Philadelphia for USAir. To derive the measure of *cost* for the medium airlines (MA), we take the minimum among the distances that we compute for Continental, USAir, America West, and Northwest. Southwest does not really have major hubs; it uses several airports, among which we consider Chicago Midway, Baltimore, Las Vegas, Houston Hobby, Phoenix, and Kansas City. With the exception of ATA, low cost carriers do not have hubs in the same sense that we mean for the largest carriers. To construct a measure of the cost, we compute the (minimum) distance from airports where LCCs had a meaningful presence. The full list of these airports is available from the authors.

Details on the Wright Amendment: The Wright Amendment restricted flight to states neighboring Texas by allowing flights with only a small commuter plane with up to a total capacity of 56 passengers. To understand how the amendment affected competition in markets out of Dallas Love, it is essential to know that one characteristic that distinguishes Southwest Airlines from other national carriers is Southwest's reliance on only one aircraft type, the Boeing 737. Southwest flies a single type of aircraft to simplify operations in terms of maintenance (older planes can be used for replacement parts), staffing, and training. Boeing 737s have a capacity of no less than 100 passengers.

The two main arguments in support of the Wright Amendment were that the amendment only applied to Love Field—not to Southwest—and that Southwest could fly nationwide from Dallas/Fort Worth, which is done by other low cost carriers.

Southwest, however, claimed that providing service at Dallas/Fort Worth would split their operation unnecessarily between the two airports, breaking their network and driving their costs up. Southwest lobbied for repeal of the Wright Amendment, claiming that it was "protectionist, anti-competitive, and anti-consumer."[5] Finally, in October 2006, a bill was enacted that determined the full repeal of the Wright Amendment in 2014.

S2.2. *Discretization*

To use the results in CHT, one needs to discretize the regressors since, currently, methods do not exist for inference in conditional moment inequalities with continuous regressors. There are many ways to discretize. We have run all our results with a coarse grid that discretizes all continuous variables into four separate bins, according to the 0, 25, 50, and 75th quantiles. In the paper, we use a more refined discretization where continuous variables are binned into at most 12 bins. The policy results were almost identical with the two types of discretization; the coefficients were slightly different.

[5]From the statement regarding repeal of the Wright Amendment from Southwest Airlines' CEO Gary Kelly, available from Southwest's web site.

REFERENCES

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2002): "Parameter Set Inference in a Class of Econometric Models," Working Paper, MIT. [4]

POLITIS, D., J. ROMANO, AND M. WOLF (1999): *Subsampling. Springer Series in Statistics*. New York: Springer. [4]

Dept. of Economics, University of Virginia, Monroe Hall, Charlottesville, VA 22903, U.S.A.; ciliberto@virginia.edu

and

Dept. of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, U.S.A.; tamer@northwestern.edu.

Manuscript received August, 2004; final revision received February, 2009.