Appendix: Market Structure and Competition in Airline Markets

Federico Ciliberto[*] Charles Murry[†] Elie Tamer[‡]

May 10, 2021

A Identification Details

This section closely follows Ciliberto and Tamer (2009) (henceforth CT), and we refer to that paper for further reading.

We provide a set of sufficient conditions that guarantee point identification of the model parameters in equation (1) in the main text. These conditions are natural in this context and rely on large support regressors. Our inference methods do not require that these conditions be satisfied as the moment inequalities adapt to partial identification, but we give them here to give intuition as to what exogenous variation might be helpful for gaining identification.

Theorem 1 *Suppose $\mathbf{Z} = (z_1, z_2)$ is such that $z_1 | z_2, \mathbf{X}$ has continuous support over the real line and that $\gamma \neq 0$. In addition, assume that $E\big([X_i \vdots X_{3-i}][X_i \vdots V_i]' \mid z_i\big)$ has full column rank for $i = 1, 2$. Suppose that there is Nash equilibrium play (possibly in mixed strategies) and that $(\nu_1, \nu_2, \xi_1, \xi_2) \perp (\mathbf{X}, \mathbf{Z})$. Then,*

1. *The parameters of the first two inequalities in (1) are identified as $z_1, z_2 \to \infty$.*
2. *In addition, $(\beta, \alpha_1, \alpha_2)$ are also point identified as $z_1, z_2 \to \infty$.*

The intuition for the above result is simple. Large support conditions are sufficient for point identification of the entry model (see Tamer, 2003). Now, for the outcome equation, we can do 2SLS *at infinity* as follows. For large values of z_1 (large negative or positive values depend on the sign of γ which can be learned fast by looking at whether large positive values of z_1 say correspond to higher likelihood of seeing a player 1 in the market) for example, player 1 is in the market with probability 1. Hence, we can use $\mathbf{X}_2$ as an instrument for V_1

[*]University of Virginia, ciliberto@virginia.edu.
[†]Boston College, charles.murry@bc.edu.
[‡]Harvard University, elietamer@fas.harvard.edu.

and do 2sls on the first outcome equation conditional on the event that $z_1 \to \infty$. Driving player 1 to enter with probability 1 eliminates the correlation between ξ_1 and $y_1 = 1$ which allows us to use "standard methods" to estimate the first outcome equation. These methods would be based on the moment condition

$$E[(X_1', X_2')'\xi_1 | z_1 \to \infty] = 0$$

Hence, what is needed for the identification of outcome equation 1 for example (arguments for the second outcome equation are similar) is two excluded variables: a standard instrument X_2 and an excluded variable from the outcome equation, z_1 in this case, that takes large values and can influence the entry of player 1. Such a variable can be one that affects fixed costs only, but not variable costs and can be exogenously moved. In the standard case, the only needed condition is an instrument X_2. So, to control for the first stage, we are required to have another instrument that can take large values. Note that the identification results in the Theorem above do NOT require that 1) the joint distribution of the unobservables be known, but requires that those be independent of the exogenous regressors, and 2) that the players play pure strategies (also here, the results in the Theorem do not require that the sign of the Δ's be known but we maintain here that the sign of these is strictly negative). On the negative side, these point identification results based on large supports lead to slow rates of convergence which makes it hard to be used with standard data sets.

Without such large support conditions, it is unclear whether we get point identification and hence it is crucial that any inference methods used is robust to failure of point identification. Basing our inference on the derived *moment inequalities* does not require that the parameter is point identified. The confidence regions that these methods use are based on inverting test statistics like the following ones.

So, under the null that $\theta = \theta^*$, we have

$$H_0: \quad E[\mathbf{G}(\theta^*, S_1 y_1, S_2 y_2, V_1 y_1, V_2 y_2, y_1, y_2) | \mathbf{Z}, X] \leq 0 \quad \text{for all} \, (\mathbf{X}, \mathbf{Z}, t_1, t_2)$$

The next theorem provides the objective function that we use to define our test statistic.

Theorem 2 *Suppose the above parametric assumptions in model (1) are maintained. In addition, assume that* $(\mathbf{X}, \mathbf{Z}) \perp (\xi_1, \xi_2, \nu_2, \nu_2)$ *where the latter is normally distributed with mean zero and covariance matrix* Σ*. Then given a large data set on* $(y_1, y_2, S_1 y_1, V_1 y_1, S_2 y_2, V_2 y_2, \mathbf{X}, \mathbf{Z})$ *the true parameter vector* $\theta = (\delta_1, \delta_2, \alpha_1, \alpha_2, \beta, \gamma, \Sigma)$ *minimizes the nonnegative objective function below to zero:*

$$Q(\theta) = 0 = \int W(\mathbf{X}, \mathbf{Z}) \| \mathbf{G}(\theta, S_1 y_1, S_2 y_2, V_1 y_1, V_2 y_2, y_1, y_2) | \mathbf{Z}, X] \|_+ dF_{\mathbf{X}, \mathbf{Z}} \qquad \text{(A.1)}$$

for a strictly positive weight function $(\mathbf{X}, \mathbf{Z})$.

The above is a standard conditional moment inequality model where we employ discrete valued variables in the conditioning set along with a finite (and small) set of t's.[1]

[1] It is possible to use recent advances in inference methods in moment inequality models with a continuum

B Computational Guide

The estimation algorithm compares moments from the data (which themselves depend on parameters) to model predicted analogues. The entry model generally predicts multiple equilibria, so for a given set of parameters and a given draw from the joint distribution of model errors, there may be multiple predictions from the model. Below, we detail the steps we use to estimate the model: (i) Evaluating the moments from the data, (ii) simulating the bounds on the moments from the model, (iii) comparing the data to the model, (iv) computation of optimization and inference, and (v) counterfactuals.

(i) Moments from the data

For a given guess of the parameters $\Theta^0 = (\alpha^0, \beta^0, \varphi^0, \gamma^0, \Sigma^0)$ we estimate the probability distribution functions for the residuals from the demand and supply equations. In the data, each market has an observed market structure, $\hat{e}_m \in E$. For all markets and all active carriers, we compute the following two residuals:

$$\hat{\xi}^{\hat{e}_m}_{jm} = log(s_{jm}) - log(s_0) - X'\beta^0 - \alpha^0 p_{jm} - \lambda ln(s_{jm|g}) \tag{A.2}$$

$$\hat{\eta}^{\hat{e}_m}_{jm} = ln(p_{jm} - [\frac{1-\lambda}{\alpha(1-\lambda s_{j|g} - (1-\lambda)s_j)}]) - \varphi W_{jm}, \tag{A.3}$$

where we are clear that the residuals are specific to a particular market structure, $\hat{e}_m$, j indexes carriers, and m indexes markets.

The moments we use in estimation are the joint distribution of these residuals. In practice, we compute the joint probability distribution (joint between supply, demand, and across all firms) function,

$$\Pr(\hat{\xi}^{\hat{e}} <= \mathbf{t_D}, \hat{\eta}^{\hat{e}} <= \mathbf{t_S} \mid \mathbf{X}, \mathbf{W}, \mathbf{Z}) \tag{A.4}$$

by constructing a histogram by binning up the domain of the residuals and counting the frequency of residuals in each bin.[2] As we describe next, the moments are constructed by taking differences between the bin-counts of the distribution of residuals from the data with the distribution of selected errors predicted by the model.

---

of moments, but these again will present computational difficulties especially in the empirical model we consider below. We detail in the next Section the exact computational steps that we use to ensure well behavior (and correct coverage) of our procedures.

[2]We estimate the CDF using histograms (using Matlab's HISTCOUNTS function). This takes much less computer memory (at least the way we are thinking about the problem). The dimensionality of the array that defines the histogram can be as small as 2-dimensional (a matrix) for a market with a single entrant and as large as 12-dimensional array for a market with six entrants (each firm has a demand and supply residual).

(ii) Model predictions

We construct the distribution of structural errors predicted to be selected by the model using simulation. For the same guess of parameters, Θ^0, we make 100 draws[3] from the joint distribution of demand, marginal cost, and fixed cost errors, $\begin{pmatrix} \xi \\ \eta \\ \nu \end{pmatrix} \sim MVN(0, \Sigma^0)$ for every market. We solve for all possible equilibria in each market for each simulation draw[4]. Solving for all possible equilibria involves finding a vector of profits (defined by the three vector-valued equations in Equation 16 in the main text) that are consistent with pure strategy Nash behavior for *every potential market structure*. Finding a vector of profits involves finding the Nash Equilibrium of prices for any particular market structure, which is itself the solution to a system of implicit non-linear equations in prices defined by the pricing first-order-conditions.[5] Because we have up to six potential entrants, we have up to 2^6 market structures to solve for profits for every simulation draw in each market[6].

When there are no pure-strategy equilibria in the entry game, we know that there exists at least one equilibrium in mixed-strategies. In that case, which happens *very* rarely in our empirical analysis, we proceed as follows. First, we determine the firms for which it is a dominant strategy not to enter. Then, we know that there will be at least one mixed strategy equilibrium where one of the remaining firms assigns a positive probability to the entry decision. Finally, we count this observation-simulation as contributing to the upper bound of the CDF of the simulated errors for all those firms.[7]

We collect all of the simulated errors that are part of equilibrium play. For example, if in one market for one simulation AA and DL are the active firms, the structural errors associated with those two carriers are the model selected errors, $(\xi^{e^*}_{AA,mr}, \xi^{e^*}_{DL,mr}, \eta^{e^*}_{AA,mr}, \eta^{e^*}_{DL,mr})$, where r indexes simulation draws. To construct the joint distribution of selected errors, we follow the same procedure of binning up the domain of the errors (using the same bin cutoffs) as we used for the computation of the distribution of the residuals. However, because of multiple equilibria, there will be an upper limit to the distribution and a lower limit. The upper limit is defined when we do not include any errors from those simulation-markets with multiple

[3]Ideally, we would like to make a very large number of draws but we restrict ourselves to 100 for computational simplicity.

[4]In practice, multiple equilibira occur less than 1% of the time at and near our estimated parameters.

[5]We employ a multi-method strategy for finding equilibrium prices. For a vast majority of the cases, iterating on the markup equation solves for the vector of equilibrium prices quickly. However, we also employ quasi-Newton root-finding when function iteration fails or moves slowly.

[6]We parallelize our code across markets. The 1 percentile of time that it takes to solve everything one time for all simulations for all markets is 8 seconds; the 99 percentile that it takes is 53 seconds. The median time that it takes is 11 seconds. At UVA, we run 28 cores in parallel at a time. We have found that the minimization is faster if we consider many starting values for a shorter period of time than if we consider fewer starting values for a longer period of time, and this motivates our use of 28 cores rather than more.

[7]For example, suppose that Firm 1 and Firm 2 are the only firms in a market, for a given simulation, for which entry is not a dominated strategy. Then, we maintain that the simulated errors for those two firms, for that simulation in that market, contribute to the upper bound of the CDF.

equilibria, as we are agnostic about equilibrium selection. The lower limit includes all errors from markets with multiple equilibria. This procedure is analogous to the procedure in CT, and yields the upper and lower joint distributions of model selected errors:

$$\Pr\left(\xi_r^e <= \mathbf{t_D}, \eta_r^e <= \mathbf{t_S}; \begin{pmatrix}\xi\\ \eta\\ \nu\end{pmatrix} \in A_e^M | \mathbf{X}, \mathbf{W}, \mathbf{Z}\right) \tag{A.5}$$

and

$$\Pr\left(\xi_r^e <= \mathbf{t_D}, \eta_r^e <= \mathbf{t_S}; \begin{pmatrix}\xi\\ \eta\\ \nu\end{pmatrix} \in A_e^U | \mathbf{X}, \mathbf{W}, \mathbf{Z}\right), \tag{A.6}$$

where (from Section 2) $\begin{pmatrix}\xi\\ \eta\\ \nu\end{pmatrix} \in A_e^U$ denotes the realizations of the errors that imply unique equilibria and $\begin{pmatrix}\xi\\ \eta\\ \nu\end{pmatrix} \in A_e^M$ denotes realization of the fixed cost error that imply multiple equilibria. To construct the density function from the simulations, we take a simple average over the psuedo-random Monte Carlo draws, indexed by r.

(iii) Constructing Moments

The moments we use for estimation involve comparing the distribution of residuals to the upper and lower bounds of the distribution of model selected errors. We take the squared difference of the bin counts that define the conditional joint distribution of residuals and model selected errors. We sum these squared differences across bin counts and across the different conditional distributions. Because of multiple equilibria, we only penalize the function if the c.d.f. of the residuals is greater than the upper limit for the selected errors or less than the lower limit for the selected errors. Notice that there are essentially two types of ways the model will not fit the data: (1) conditional on a market structure and (X, Z, W), the residuals have a different distribution than the selected errors, and (2) the model predicts different market structures than the data.

To choose the t's in the grid, we proceed in two steps. First, we determine the distributions of the demand and marginal cost residuals when we estimate the model with GMM without selection. We use this to learn over what support the residuals are defined, in terms of their max and min value. Using this approach, we selected the following values for the t of the demand: [-10;-7.5;-5;-2.5;0;2.5;5;7.5;10]. And we chose the following values for the t of the marginal cost: [-2;-1.5;-1;-0.5;0;0.5;1;1.5;2], or one-fifth the scale of the demand errors. In our experimentation when estimating Column 1 of Table 4 (the exogenous case) we found that this proportional relationship is important. Ideally, one would want to have a very fine grid for the t but there is a trade-off because of memory limitations (explained above) associated with storing so many cells used to construct the histogram estimate.

(iv) Objective Function Minimization, Computational Details, and Inference

The minimization of the distance function given by Equation (A.1) in the main text and described above is computationally intensive because we have to use simulation methods to integrate two multi-dimensional distribution functions and then compare them. In addition to constructing the distribution functions, we need to solve for Nash equilibria in many markets, and for many possible combinations of firms in each market. We need to do these things many times because the objective function may be non-smooth and non-convex, so finding a set of parameters that minimize the objective function may be taxing.

The keys to finding the global minimum are:

1. Parallel computing;
2. Good initial guesses on as many parameters as possible;
3. Using many different starting values;
4. Using flexible minimization routines that mix different built-in algorithms.

Each one of these ingredients is important in finding a global minimum. Overall, we reach the area of the global minimum in approximately three days and the optimization is completed in approximately seven days. In practice, we have starting values for many of the parameters from IV regressions that do not account for endogenous entry.

For inference purposes, we continue the minimization longer in order to collect as many parameters close to the *argmin*. We use Matlab's optimization algorithms to sample the objective function and we save the results to get a snapshot of the surface of the function. In addition, we randomly and non-randomly sample parameters close to the minimum to achieve good coverage around the minimum in order to construct confidence regions.

Good Initial Guess on as many parameters as possible: The GMM estimation that assumes exogenous market structure provides us with natural starting values for the parameters of the utility and marginal cost functions. In addition to the GMM estimated parameter values, we also consider the parameters that are estimated when we run the exogenous entry model using our methodology.

To get starting values for the parameters in the fixed cost function, and for the remaining parameters in the variance covariance matrix, we choose values that are small (for example, between -2 and 2, in the case of the variance of the fixed costs). We expand the bounds on these parameters as we continue our search.

Many Starting Values: We start with multiple initial values, which are derived as follows. For each of the initial guesses above, we find a reasonable interval around those guesses in the sense that the intervals are on the same scale as the standard errors from the GMM estimation and the interval implies sensible economic predictions, for example positive marginal costs and markups that are not near zero. For example, for the price parameter estimate, which is equal to -0.0229 in the GMM, we prepare an interval equal to [-0.035, -0.015]. We repeat this exercise for all the parameters. An important remark: Recall that in order to limit the space over which to draw for the argmin, we have standardized all the exogenous variables.

Next, we draw up to *50,000* independent random draws from these intervals. Out of these 50,000 starting values, we select the 10 that are associated with the lowest distance function values. This first step takes approximately one day of time, but we save these function evaluations, so this is itself part of the minimization and confidence function construction processes. Our next step is to use algorithms in the Global Optimization Toolbox in Matlab to minimize the function starting from the first round of 10 lowest values.

Multiple Iterations of Flexible Minimization Routines: In our experimentation we have used different combinations of three algorithms in the Global Optimization Toolbox in Matlab: simulannealbnd, patternsearch, and fminsearchOS. We have found that patternsearch provides the best minimization results after we draw the 50,000 initial values, as described above. Therefore, we take the best 10 values out of the 50,000, and run patternsearch.

We have found that after 24 hours, patternsearch converges to a new parameter value which usually depends on the starting values, and that is why it is crucial to draw as many starting values as we do. This is because the distance function is highly nonlinear and the minimization problem is complex.

At this point we take the 10 local minima after running patternsearch, and reiterate the process described in the Section above (Many Starting Value), but adapting the bounds. We draw up to 50,000 independent random draws. We run patternsearch again on the 10 that are associated with the lowest distance function.

In our work, we have run few iterations of this two-step process of i) drawing randomly, then ii) using patternsearch this process this time. We have determined that this process is the one that reaches the global minimum in the most efficient way. The step of continuing to draw randomly close to the current minimum in the minimization search is crucial.

We finish the estimation with fminsearchOS, a more flexible implementation of Matlab's fminsearch, which can be found on Matlab's FileExchange platform. There are no bounds on the parameters when we run fminsearchOS. fminsearchOS only takes few hours to converge.

Overall, for the estimation of Column 3 in Table 4 of the paper we end up with 614,068 parameter guesses. The minimization process takes approximately one week of time. Clearly, as new and more powerful minimization algorithms are introduced, it should be possible to find the global minimum in a shorter time.

Plots Around the Minimum Following a Referee's suggestion, we show plots of the objective function around the minimum as we vary each parameter, one by one. The plots are in Figure A1. Although this is not a formal proof of optimality, the plots lend credibility to the outcome of our estimation routine.[8]

[8] The domain of each plot is four times our confidence interval. The shaded region corresponds to our confidence region.

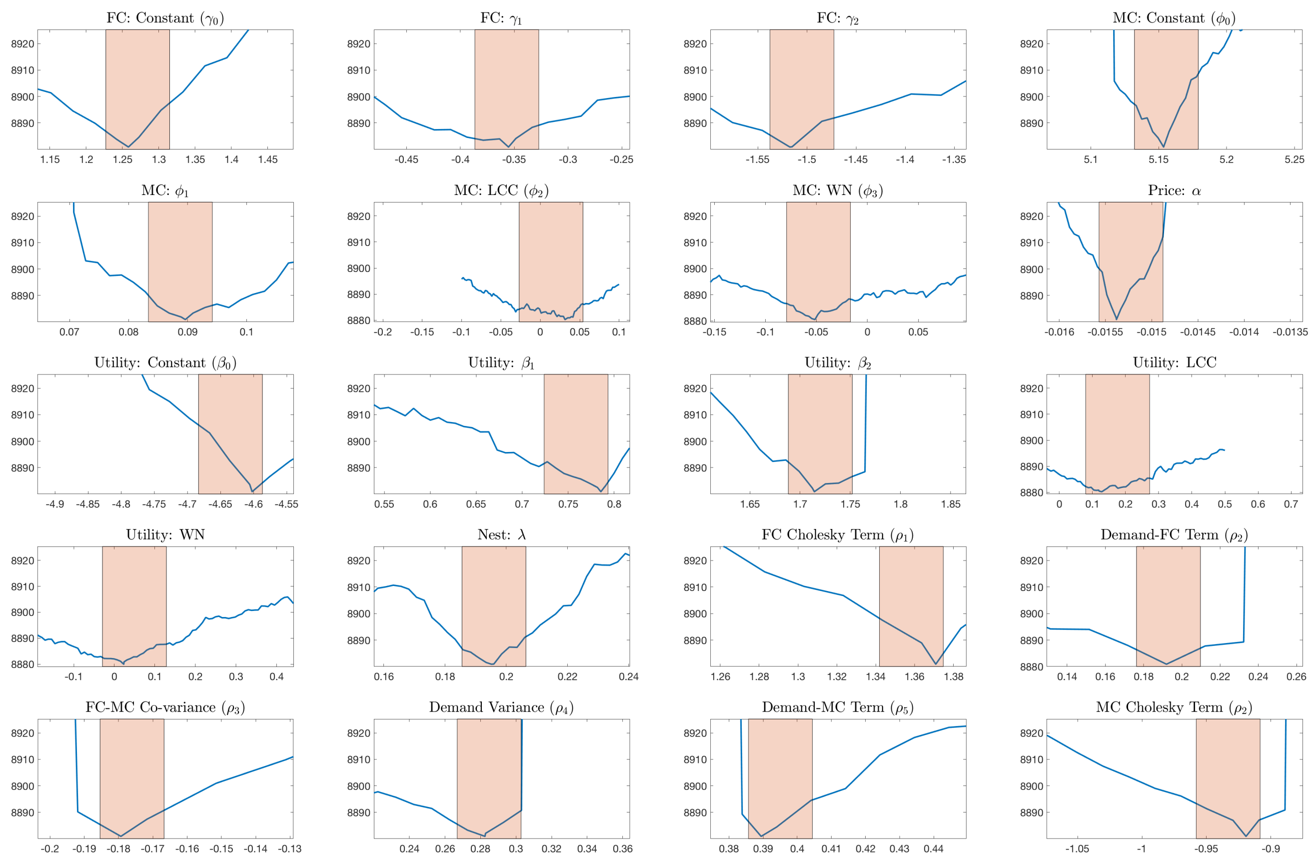


Figure A1: Objective Function in Each Parameter Dimension

Inference The construction of confidence regions follows Chernozhukov, Hong, and Tamer (2007), which involves obtaining critical values via subsampling. In practice, we use 100 subsamples of one-quarter the size of the original dataset and start the subsampling routine from the *argmin* we found in the initial estimation. We compute the confidence region for the point using the procedure outlined in the On-line Appendix of CT, page 5.

Note on normalizing the objective function $Q(\theta)$: after finding a minimizer of the objective function, and so obtaining a value $\eta_n = \min_\theta Q_n(\theta)$ we normalize the *sample* objective function and use $Q_n^*(\theta) = Q_n(\theta) - \eta_n$ to construct confidence regions (as suggested in CHT). This is similar to constructing constructing χ^2 confidence regions in GMM and plays a role in that it guarantees that in finite samples, the confidence intervals are non-empty. Under a well specified model, this normalization plays no role asymptotically. If the model is misspecified (i.e., $\min Q(\theta) > 0$), then this normalization will guarantee that we are constructing confidence regions for the $arg\min$ of $Q(\theta)$.

(v) Details of Counterfactuals

We predict the effects of a AA-US merger for four different assumptions about the new merged firms:

1. the surviving firm, AA, takes the best observed and unobserved characteristics from the pre-merger AA and US,

2. the surviving firm inherits the best observed characteristics, but we take AA's unobservable,

3. the surviving firm inherits the best observed characteristics, but we re-draw errors for the firm,

4. the surviving firm takes the mean values of the observed and unobserved characteristics from the pre-merger AA and US, and

To construct confidence intervals for each of these counterfactual scenarios, we draw 100 parameter vectors from the confidence set and evaluate the counterfactual equilibrium market outcomes for each parameter vector. For a specific table, for example Table 8, starting from all of the counterfactual simulations we condition on a particular pre-merger market structure. For a particular parameter vector, we compute the upper and lower bounds for the number of times we observe the market structure changing to each possible market structures, post merger. So for Table 8, we took all the simulations that (for a particular parameter vector) predicted AA/US Duopoly before the merger and then computed then counted the number of times we observed an AA monopoly post-merger. Then, to get the 95% confidence sets, we take the 2.5 percentile and 97.5 percentile of the probability of observing each market structure, across the 100 parameter vectors we drew from the original confidence set. We follow the same procedure for the prices – always conditioning on the pre-merger market structure.

(vi) Timing and Acknowledgments

At the beginning of our computational work, in 2011, we ran testing for the code on multiple systems, including the XSEDE resources Gordon and Trestles at the San Diego Supercomputing Center. Performance and scaling tests on Gordon indicated at most 32 workers (cores) provided the shortest execution time before communication overhead to the workers becomes significant. The computationally intense estimation of our models at the time in a relatively short period of time was made feasible because of the use of XSEDE resources.[9] In our experimentation we found that having 200 or 500 simulations did not make a difference in our results, while the time taken was much larger. Thus, the costs of more simulations outweighed the benefits.

More recently, we have used two other resources to implement the optimization routine, the HPC system at the University of Virginia known as Rivanna, and the HPC system at Penn State University known as ACI. Rivanna is a 4800-core, high-speed interconnect cluster, with 1.4 PBs of storage available in a fast Lustre filesystem. ACI is a 23,000 core high-speed cluster with 20 PBs of storage and 640 teraflops of total peak performance.

We gratefully acknowledge the use of both the XSEDE resources and those at the University of Virginia and Penn State University.

C Data Construction

The main data are from the domestic *Origin and Destination Survey (DB1B)*, the *Form 41 Traffic T-100 Domestic Segment (U.S. Carriers)*, and the *Aviation Support Tables : Carrier Decode*, all available from the Department of Transportation's National Transportation Library. We also use the US Census for the demographic data, specifically to get the total population in each Metropolitan Statistical Area. The *Origin and Destination Survey (DB1B)* is a 10 percent sample of airline tickets from reporting carriers. The dataset includes information on the origin, destination, and other itinerary details of passengers transported, most importantly the fare. The *Form 41 Traffic T-100 Domestic Segment (U.S. Carriers)* contains domestic non-stop segment data reported by US carriers, including carrier, origin, destination of the trip. The dataset *Aviation Support Tables : Carrier Decode* is used to clean the information on carriers, more specifically to determine which carriers exit the industry over time, and which one merge, or are owned by another carrier.

We define a market as a unidirectional trip between two airports, irrespective of intermediate transfer points. For example, we will assume that the nonstop service between Chicago O'Hare (ORD) and New York La Guardia (LGA) is in the same market as the connecting service through Cleveland (CLE) from ORD to LGA. The market ORDLGA is a different market from LGAORD.

[9]John Towns, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D. Peterson, Ralph Roskies, J. Ray Scott, Nancy Wilkens-Diehr, "XSEDE: Accelerating Scientific Discovery", Computing in Science & Engineering, vol.16, no. 5, pp. 62-74, Sept.-Oct. 2014.

We follow Borenstein (1989) and assume that flights to different airports in the same metropolitan area are in separate markets. To select the markets, we merge this dataset with demographic information on population from the U.S. Census Bureau for all the Metropolitan Statistical Areas of the United States. We then construct a ranking of airports by the MSA's market size. Our final dataset includes a sample of markets between the top 100 Metropolitan Statistical Areas, ranked by the population size in 2012. We exclude the Youngstown-Warren Regional Airport, Toledo Express Airport, St. Pete-Clearwater International Airport, Muskegon County Airport, and Lansing Capital Region International Airport because there are too few markets between these airports and the remaining airports.

Then, we proceed to further clean the data as follows. We drop: 1) Tickets with more than 6 coupons overall, or more than 3 coupons in either direction if a round-trip ticket; 2) Tickets involving US-nonreporting carrier flying within North America (small airlines serving big airlines) and foreign carrier flying between two US points; 3) Tickets that are part of international travel; 4) Tickets involving non-contiguous domestic travel (Hawaii, Alaska, and Territories) as these flights are subsidized by the US mail service; 5) Tickets whose fare credibility is questioned by the DOT or for which the bulk fare indicator was equal to 1 ; 6) Tickets that are neither one-way nor round-trip travel; 7) Tickets including travel on more than one airline on a directional trip (known as interline tickets), here identified by whether there was a change in the ticket carrier for the ticket.

Next, we follow the approach in Borenstein (1989) and Ciliberto and Williams (2014) and consider a round-trip ticket as two directional trips on the market, and the fare paid on each directional trip is equal to half of the round-trip fare. A one-way ticket is one directional trip.

Moreover, as in Berry and Jia (2010) and Ciliberto and Williams (2014), tickets sold under a code-share agreement (for example, a ticket sold by USAir on a United operated flight) are allocated to the airlines that sold the tickets (so, in the example, to USAir). This is consistent with the notion that the ticketing carrier has access to the "metal" (the seats) of the operating carrier. Notice that this implies that there can be observations where the airline does not have any nonstop routes out of an airport, but the airline can sell tickets for flights out of that airport.

We then drop: 1) Tickets with fares less than 20 dollars; 2) Tickets in the top and bottom one percentiles of the year-quarter fare distribution, and tickets for which the fare per mile (the yield) was in the top and bottom one percentiles of the year-quarter yield distribution.

We then aggregate the ticket data by ticketing carrier and thus the unit of observation is market-carrier-year-quarter specific.

Next, we drop markets whose distance is less than 150 miles. We also drop airlines that served fewer than 90 passengers in a quarter. Finally, we determine the markets that are not served by any airline, but that could be potentially served by one. These are the markets that were served at least 80 percent of all quarters between the first quarter in 1994 and the first quarter in 2017.

The airlines in the initial dataset are: American, Alaska, JetBlue, Delta, Frontier, Allegiant, Spirit, Sun Country, United, USAir, Virgin, Southwest. By the second quarter of

2012, Southwest had completed the acquisition of AirTran, although the two carriers were still issuing tickets with different code (FL vs WN). As in Ciliberto and Tamer (2009), we deal with how to treat regional airlines that operate through code-sharing with national airlines as follows. 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.

The low cost type is composed of: Alaska, JetBlue, Frontier, Allegiant, Spirit, Sun Country, Virgin. We re-elaborate their data as follows. The LCC's number of passengers is the sum of the passengers over all the LCCs that serve a market. The LCC's price is the passenger weighted mean of the prices charged by all the LCC airlines in a market. For the explanatory variables we take the maximum value among the low cost carriers serving a market of the variables *Origin Presence*, *Destination Presence*, *Nonstop Network Origin*, *Nonstop Network Destination*. We also take the maximum of the categorical variables that indicate whether a firm is a potential entrant in a market.

After this preliminary cleaning, we compute the 95 percentile of the mean prices and yield per mile, and we drop markets where prices and yields above these values were observed.

In order to compute the confidence intervals as in Chernozhukov, Hong, and Tamer (2007) we discretize the exogenous variables. The discretization is done as follows. First, we standardize the continuous variables. Then, we construct intervals where the thresholds are given by -1, -0.5, 0, 0.5, 1, as well as integers such as -2, 2, -3, 3. The discretization affects the variables in both the solving the model as well as the values of the instruments. When also estimate the exogenous-entry GMM specification with these discretized variables.

D Robustness Analysis

This Section investigates how demand estimates change with changes in the modeling of the demand and in the nature of the exogenous variation that identifies the demand coefficients. The results are before the discretization of the variables.

Column 1 of Table A2 presents the baseline results from running a standard OLS regression. The price coefficient is estimated equal to -0.004, and it implies a median elasticity of -0.902, which is inconsistent with a model of profit maximization. There are 15,100 observations out of 22,445 for which the elasticity is larger than -1.

Column 2 of Table A2 presents the baseline results from running a standard two stage least squares nested logit regression, when we use *Nonstop Destination* and *Nonstop Origin* as instrumental variables. We use both the values of the firm associated with the observation as well as the values of the potential competitors. This is analogous to the identification strategy in Bresnahan (1987). The coefficient estimate of the price is now equal to -0.012, and the median elasticity is -3.005.

Column 3 presents the results when we estimate a nested logit as in Berry (1994) using the same instrumental variables that we used in Column 2. We find the coefficient of price equal to -0.028, the coefficient of the nesting parameter equal to 0.528, and the corresponding

Table A2: *Parameter Estimates with Exogenous Market Structure*

	OLS Logit	Simple Logit IV	Nested Logit IV	Nested Logit IV	Nested Logit IV
Demand					
Price	-0.004	-0.012	-0.028	-0.020	-0.026
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
σ	–	–	0.529	0.361	0.420
	–	–	(0.016)	(0.012)	(0.016)
Distance	-0.241	0.161	0.923	0.518	0.790
	(0.016)	(0.027)	(0.038)	(0.028)	(0.037)
Origin Presence	0.007	0.006	0.004	0.022	0.005
	(0.000)	(0.000)	(0.000)	(0.048)	(0.000)
LCC	1.004	0.458	-0.505	0.082	-0.338
	(0.037)	(0.049)	(0.062)	(0.041)	(0.060)
WN	1.201	0.957	0.062	0.445	0.232
	(0.022)	(0.027)	(0.040)	(0.031)	(0.039)
Constant	-8.148	-6.338	-1.936	-4.066	-2.736
	(0.046)	(0.108)	(0.179)	(0.127)	(0.175)
Elasticities and Percentage Contribution Margins					
Median Elasticity	-0.901	-3.005	-11.975	-6.578	-9.1220
Elasticities $>= -1$	15,100	7	0	0	0
Nonstop Destination IV	No	Yes	Yes	Yes	Yes
Nonstop Origin IV	No	Yes	Yes	Yes	No
Potential Entrants IV	No	No	No	Yes	No

median elasticity equal to -11.975.

Column 4 presents the results when we include the information on the potential entrants as instrumental variables. In practice, we add six variables as instrumental variables, one for each of the six firms (AA, DL, UA, LCC, WN, US). The coefficient estimate of price is now -0.020, and the nesting parameter is estimated equal to 0.361. These values are very similar to those in our GMM estimates in Table 4 of the paper. The corresponding median elasticity is equal to -6.578.

Finally, Column 5 of Table A2 shows the results if we maintain that only *Nonstop Destination* can be used as instrumental variables. This is a key mantained assumption in the identification strategy in Berry and Jia (2010). We find now that the price coefficient is estimated equal to -0.026 and the nesting parameter is 0.420. The corresponding median elasticity is -9.122.

Overall, Table A2 shows that the parameter estimates of the price coefficients are stable across Columns 3-5, and show that the information on the potential entrants, as well as the inclusion of *Nonstop Origin* as instrumental variables delivers estimates of the median elasticity that are closer to previous work. More specifically, Berry and Jia (2010) use data from 1996 to 2006 and estimate it between -2 and -3 in 2006 and trending upward from 1999.

Ciliberto and Williams (2014) use data from 2006 to 2008 and estimate the aggregate price elasticity to be equal to -4.320 in their model that does not allow for collusive behavior. Berry and Jia (2010) and, later on, Ciliberto and Williams (2014), use a two-type model of demand, where they distinguish between two types, a coach type whose elasticity both papers estimate to be between -6 and -6.5; and a business type, whose elasticity of demand both papers estimate to be around -0.5. The estimates of the aggregate price elasticity differs in Berry and Jia (2010) and Ciliberto and Williams (2014) because they estimate different fractions of coach and business travelers. Berry and Jia (2010) estimate the share of business passengers between 41 and 49 percent. Ciliberto and Williams estimate the share of business passengers to be 34 percent. We infer that the average elasticity of demand doubled between the period analysed by Berry and Jia and the one analysed by Ciliberto and Williams, because of an increase in the share of economy passengers. Our dataset is from 2012, well after the ones used by Berry and Jia (2010) and by Ciliberto and Williams (2014), and therefore the increase in price elasticity is consistent with an increase in the share of economy passengers.

E Numerical Exercise

We run a series of numerical exercises to show that GMM estimates of markups are biased if the true model has endogenous entry and to show that our estimation methodology works well when we know the true parameters.

First, we present a slightly simplified version of our model. The simplifications include fewer demand and cost variables. In practice, the model includes the minimum number of parameters to make it comparable to our empirical analysis. The model is represented by the following system of conditions:

$$\begin{aligned}
&Demand: && ln(s_{jm}) = \alpha p_{jm} + c_1 + X_{jm} b_1 + \lambda ln(s_{jgm}) + \xi_{jm} && \text{(A.7)}\\
&Supply: && ln(c_{jm}) = c_2 + b_2 X_{jm} + \eta_{jm} && \text{(A.8)}\\
&Entry: && y_j = 1 \Leftrightarrow \pi_j \equiv (p_{jm} - c_{jm}) M_m s_{jm} - exp(c_3 + b_4 Z + \nu) FC_{jm} \geq 0, && \text{(A.9)}\\
& && && \text{(A.10)}
\end{aligned}$$

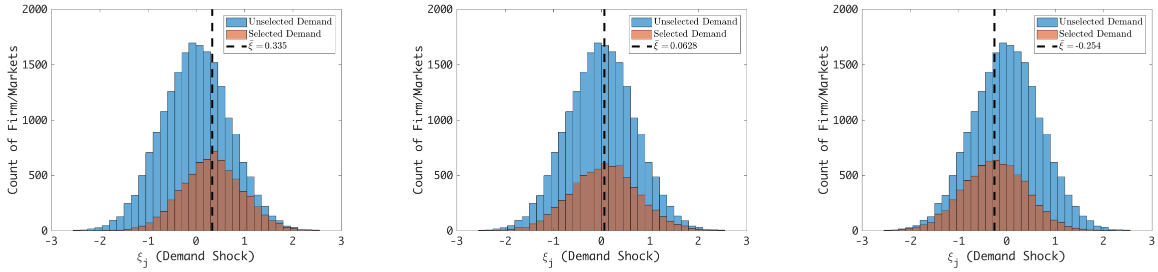
where the epxression for demand and marginal costs are the following,

$$s_{jm} = \frac{exp(\alpha p_{jm} + X_{jm}\beta + \xi_{jm})}{1 + \sum_k exp(\alpha p_{jm} + X_{km}\beta + \xi_{km})} \tag{A.11}$$

$$c_{jm} = p_{jm} - \frac{\alpha(1-\rho)}{1 - \rho s_{jgm} - (1-\rho) s_{jm}}. \tag{A.12}$$

We assume the following variance covariance matrix for a particular airline:

$$\Sigma_{jm} = \begin{bmatrix} \sigma_\xi^2 & \sigma_{\eta\nu} & \sigma_{\xi\nu} \\ \sigma_{\eta\nu} & \sigma_\eta^2 & \sigma_{\eta\nu} \\ \sigma_{\xi\nu} & \sigma_{\eta\nu} & \sigma_\nu^2 \end{bmatrix}.$$



(a) $cov(\xi, \nu) = -0.4$ (b) $cov(\xi, \nu) = 0$ (c) $cov(\xi, \nu) = 0.4$

Figure A2: Distribution of Demand Errors (ξ) For Different Covariances

As in the main text, we assume that the correlation is only among the unobservables within a firm, and not between the unobservables of the K_m firms. This specification also restricts the correlations to be the same for each firm and clearly reduces the parameters to be estimated. However, the specification is rich compared to existing methods.

We generate covariates from a standard normal, one covariate for demand and one for fixed costs. We also randomly generate market sizes for each market.

First, we can describe the role of selection in the model by displaying the distribution of errors pre-selection and post-selection for various values associated with the covariance matrix of the unobserved terms. In Figure A2 we display three graphs of histograms of errors. In each graph, the larger histogram represents the pre-selected distributions of demand errors for all of the potential firms in all of the simulated markets. This distribution is drawn from an underlying joint normal with a mean of zero and the covariance matrix parameters displayed in Column 1 of Table A3, except that we vary the correlation between demand and fixed costs. In each graph, the smaller histogram represents those demand errors from firms in markets that the model predicts to enter, or in other words, the selected errors. It is clear that the distribution of selected demand errors changes as the covariance between demand and fixed costs changes. In the model, negative correlation between demand and fixed cost shocks implies positively selected firms, which is intuitive and can be seen in the first panel of Figure A2. The middle panel shows that without any correlation, the model induces only a slightly shifted distribution of demand errors.[10] When demand and fixed cost shocks are positively correlated, the distribution of selected demand shocks is shifted to the left. We would expect the corresponding bias in elasticity estiamtes to vary based on the values of the covariance matrix as well.

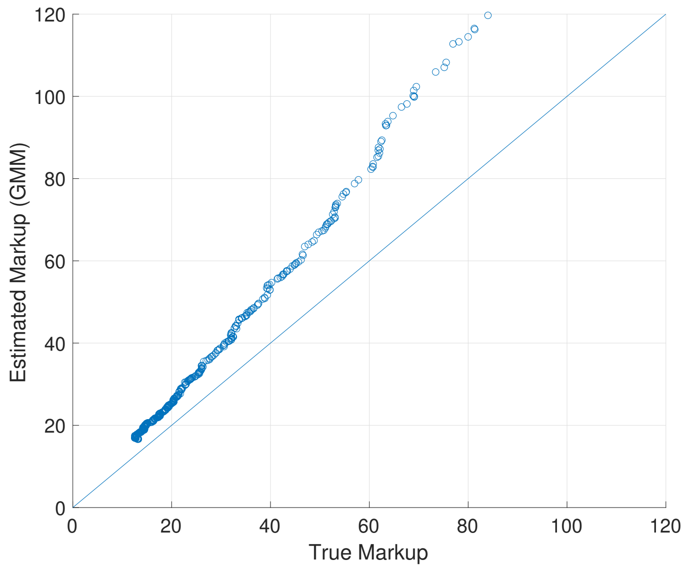
Monte Carlo Simulation: Bias in "Standard" Model

We document the bias from estimating a standard model that does not account for selection. To do this, we solve the model 1000 times for different random draws of the covariates, errors,

[10]We computed this numerical exercise for many different parameters and have found positive and negative selection in the case where $cov(\xi, \eta) = 0$. We chose to display this particular case because these are the parameters we use for the estimation exercise below.

price parameter (α), nest parameter (λ), covariance matrix, and sets of potential entrants. For each of the 1000 generated data sets, we estimate demand using the method suggested in Berry (1994) and compute the implied markups. In Figure A2, we graphically compare the implied markups from GMM to the true markups used to generate the 500 different draws of data, varying the price sensitivity across datasets as well. It is clear that the estimates are systematically different than the true values.

Figure A2: GMM Bias in Markups Across Different Parameter Values



Note: Plot of true markups versus estimated markups using GMM that does not account for endogenous selection/entry. Each point represents a different draw of data, errors, price parameter, and nesting parameter.

Model Estimation with Endogenous Entry

Next, we estimate the model using simulated data, employing the methodology we present in Sections 2 and 3 of the main text. The true parameters are in Column 1 of Table A3. In Column 2 we present the estimates using GMM, not accounting for selection a la Berry (1994). In the Column 3, we present the 95% confidence intervals using our methodology.

Our methodology does quite well. Most of the true parameters lie within their associated confidence intervals using our methodology, and in many cases the confidence intervals are tight.[11] In particular, our methodology does a much better job at estimating the price coefficient than GMM.

[11] The confidence intervals here are larger than in our empirical exercise. One reason is that our real data looks irregular in the sense that it does not look normal like the fake data – in this sense the real data might

It is not surprising that the price parameter, in particular, suffers from bias in the GMM estimation, because it links all three model conditions through its role in determining markups (and, thus, the entry profit threshold condition as well).

Table A3: Parameter Estimates Using Simulated Data

	True	GMM	Endogenous Entry
Demand			
Price	-0.02	-0.034 (0.005)	[-0.029, -0.021]
Constant	-3	0.077 (0.995)	[-3.124, -1.842]
X	0.5	1.284 (0.278)	[-0.819, 0.903]
Nest (λ)	0.30	0.307 (0.166)	[0.273, 0.360]
Marginal Cost			
Constant	5	5.067 (0.003)	[4.670, 5.272]
X	0.5	0.375 (0.003)	[0.107, 0.610]
Fixed Cost			
Constant	3	–	[1.071, 3.318]
Z	-0.5	–	[-0.594, -0.231]
Variance-Covariance			
Marg. Cost Variance	0.10	0.074	[0.181, 0.263]
Demand Variance	2	3.254	[1.345, 3.498]
Demand-FC Covariance	-0.10	–	[-0.027, 0.147]
Demand-MC Covariance	0.20	–	[0.551, 1.200]
MC-FC Correlation	0.10	0.3945	[0.286, 0.607]

Column 1: parameter values used to create simualted data. Column 2: Standard GMM estimation. Column 3: Estimation using the methodology described in Section 2. Standard errors in parentheses in Column 2. Columns 3 contain 95% confidence bounds constructed using the method in Chernozhukov, Hong, and Tamer (2007). The dataset includes 5000 markets with up to four potential entrants. We use 100 draws to simulate the joint distribution of errors. As in the empirical application, we fix the fixed cost variance at 0.5.

better satisfy large support conditions. Second, we use more bins to discretize the real data. Third, there is more variation in potential entrants in the real data.

References

Berry, Steven T. 1994. “Estimating discrete-choice models of product differentiation.” *The RAND Journal of Economics* :242–262.

Chernozhukov, Victor, Han Hong, and Elie Tamer. 2007. “Estimation and confidence regions for parameter sets in econometric models.” *Econometrica* 75 (5):1243–1284.

Ciliberto, Federico and Elie Tamer. 2009. “Market structure and multiple equilibria in airline markets.” *Econometrica* 77 (6):1791–1828.

Tamer, Elie. 2003. “Incomplete simultaneous discrete response model with multiple equilibria.” *The Review of Economic Studies* 70 (1):147–165.