Online Appendix to "Price Competition and Endogenous Product Choice in Networks: Evidence from the US Airline Industry"

Christian Bontemps* Cristina Gualdani† Kevin Remmy‡

March 2025

*Email: christian.bontemps@tse-fr.eu, ENAC & Toulouse School of Economics, University of Toulouse, Toulouse, France.

†Email: c.gualdani@qmul.ac.uk, Queen Mary University of London, London, United Kingdom.

‡Email: kevin.remmy.001@gmail.com, University of Mannheim, Mannheim, Germany.

A Proofs

Proof of Theorem 1 (i) For each inequality r, and each $\gamma \in \mathbb{R}^{\mathrm{P}}$,

$$\frac{1}{\mathrm{M}} \sum_{m=1}^{\mathrm{M}} \left(Z_{m,r} B_m^\top \gamma - Z_{m,r} A_m \right) - \mathbb{E}\left(Z_{m,r} B_m^\top \gamma - Z_{m,r} A_m \right) \xrightarrow[\mathrm{M}\to\infty]{P} 0.$$

Theorem 5.5 of Shapiro et al. (2014) shows the convergence of the support function of the estimated set $\hat{\Gamma}_I$ toward the support function of the true set Γ_I.

(ii) comes from Shapiro et al. (2014), Theorem 5.11. Shapiro's proof is based on the i.i.d. assumption but can be adapted as long as a Central Limit Theorem is valid for the set of inequality constraints.

The uniformity in q for both (i) and (ii) comes from the compactness of the unit ball. ■

Proof of Theorem 2 (i) From the convergence of $\widehat{\Gamma}_I^\kappa$ to Γ_I^κ with respect to the Hausdorff distance and the isomorphism with the support function, we have

$$\sup_{q,\ \|q\|\leq 1} |\delta(q, \Gamma_I^\kappa) - \delta(q, \Gamma_I)| = d_H(\Gamma_I^\kappa, \Gamma_I) = \mathrm{V}/\kappa. \tag{A.1}$$

Then, Theorem 5.11 of Shapiro et al. (2014) combined with the compactness of the unit ball shows that,

$$\sup_{q,\ \|q\|\leq 1} |\hat{\delta}(q, \Gamma_I^\kappa) - \delta(q, \Gamma_I^\kappa)| \xrightarrow[\mathrm{M}\to\infty]{P} 0. \tag{A.2}$$

Combining (A.1) and (A.2) with the triangular inequality leads to (i).

(ii) Theorem 5.11 of Shapiro et al. (2014), combined with the delta method for the asymptotic variance of the estimated constraints leads to the pointwise expansion. Denoting $b_r = \mathbb{E}(Z_{m,r} B_m)$ and $a_r = \mathbb{E}(Z_{m,r} A_m)$, for $r = 1, \ldots, \mathrm{R}$, we have

$$\begin{aligned}
\frac{\partial g_\kappa}{\partial b_r}(\gamma) &= \frac{\gamma \exp\left(\kappa \left[b_r^\top \gamma - a_r\right]\right)}{1 + \exp\left(\kappa \left[b_r^\top \gamma - a_r\right]\right)} = \frac{\gamma}{1 + \exp\left(-\kappa \left[b_r^\top \gamma - a_r\right]\right)}, \\
\frac{\partial g_\kappa}{\partial a_r}(\gamma) &= \frac{-1}{1 + \exp\left(-\kappa \left[b_r^\top \gamma - a_r\right]\right)}.
\end{aligned}$$

Therefore:

$$\begin{aligned}
\sqrt{\mathrm{M}}(\hat{g}_\kappa(\gamma) - g_\kappa(\gamma)) &= \sum_{r=1}^{\mathrm{R}} \frac{\sqrt{\mathrm{M}} \left((\hat{b}_r - b_r)^\top \gamma - (\hat{a}_r - a_r) \right)}{1 + \exp\left(-\kappa \left[b_r^\top \gamma - a_r\right]\right)} + o_P(1), \\
&= \sum_{r=1}^{\mathrm{R}} \frac{W_r(\gamma)}{1 + \exp\left(-\kappa \left[b_r^\top \gamma - a_r\right]\right)} + o_P(1).
\end{aligned}$$

The uniformity in q comes from the compactness of the unit ball. ■

Proof of Corollary 1 See the proof of Proposition 10, Section C.2.1 of Bontemps et al. (2012). The key argument is the uniqueness of the argmin of the test statistic combined with (ii) of Theorem 2. ■

B Existence of Nash Equilibrium Networks

As discussed in Section 3.3 of the main paper, proving the existence of a pure strategy Nash equilibrium (PSNE) $G \coloneqq (G_f : f \in \mathcal{N})$ is difficult due to the presence of spillovers from entry across markets on the demand, marginal cost and fixed cost sides.

Berry (1992) establishes the existence of a PSNE in one of the first empirical models of entry that incorporates strategic interactions between firms in the second-stage pricing game. His proof relies on the assumption that the entry decisions are independent across markets. It is therefore not applicable to our framework. Another approach used in the network formation literature to show the existence of a PSNE is to represent the model as a potential game (Monderer and Shapley, 1996). This is possible if the payoff function is additive separable in the linking decisions and linear in the spillovers (as for example in Mele, 2017), which is not the case here. Alternatively, it is possible to show the existence of a PSNE under the assumption that the game is supermodular, in order to exploit the fixed point theorem for isotone mappings (Topkis, 1979). However, supermodularity does not hold in our setting due to the second-stage competition between airlines. Finally, one could try to decompose the original game into "local" games such that the original game is in equilibrium if and only if each local game is in equilibrium (Gualdani, 2021). In turn, the existence of a PSNE in each local game—which is typically easier to establish—is sufficient for the existence of a PSNE in the original game. However, the classes of spillovers considered in our model do not allow us to implement such a decomposition.

One might also ask whether allowing for private fixed cost shocks could simplify the existence proof. Espín-Sánchez et al. (2023) prove equilibrium existence in an entry model where firms have some private information at the entry stage. However, they do not allow for multi-product firms and they do not allow for spillovers from entry across markets. Moreover, in our setting it is more reasonable to assume that the fixed cost shocks are common knowledge among airlines, as discussed in Section 3.2 of the main paper.

Note that the moment inequalities in Section 4.2 of the main paper are based on necessary conditions for PSNE. Therefore, one could consider a first-stage equilibrium notion that is weaker than PSNE. In particular, given our focus on one-link deviations, inequalities (11) and (12) resemble the notion of pairwise stability used in network theory, according to which no player has profitable deviations by adding or removing a link (Jackson and Wolinsky, 1996). Definition B.1 introduces a notion of first-stage equilibrium along the lines of pairwise stability.

Definition B.1. (*Pairwise Stability*) The networks $G_1, \ldots, G_{\mathrm{N}}$ represent a pairwise stable outcome if, for each market $\{a, b\} \in \mathcal{M}$ and airline $f \in \mathcal{N}$, it holds that

$$G_{ab,f} = 0 \Rightarrow \Delta\Pi_{(+ab),f} \leq \Delta\mathrm{FC}_{(+ab),f},$$
$$G_{ab,f} = 1 \Rightarrow \Delta\Pi_{(-ab),f} \geq \Delta\mathrm{FC}_{(-ab),f},$$

where the quantities $\Delta\Pi_{(+ab),f}$, $\Delta\Pi_{(-ab),f}$, $\Delta\mathrm{FC}_{(+ab),f}$, and $\Delta\mathrm{FC}_{(-ab),f}$ are defined in Section 4.2 of the paper. $\diamond$

Note that although pairwise stability is a weaker equilibrium notion than PSNE, establishing the existence of a pairwise stable outcome does not appear to be easier in our setting. In particular, according to Jackson and Watts (2002), for any payoff function there is either a pairwise stable outcome or a closed cycle.[1] A typical way used in the literature to exclude the presence of closed cycles is to show that the model can be represented as a potential game, as discussed by Jackson and Watts (2001) and Hellmann (2013). As before, however, this is possible if the payoff function is additive separable in the link decisions and linear in the spillovers (as in Sheng, 2020), which is not true in our case.

C How to Deal with Incoherence

In Section 4.2 of the main paper, we have constructed the identified set for the first-stage parameters under the assumption that PSNE networks exist for each parameter value and variable realization. As discussed above, proving the existence of PSNE networks is difficult. Therefore, it is legitimate to wonder whether one should modify the definition of the identified set when non-existence is possible, i.e., when our model is incoherent in the terminology of Tamer (2003) and Lewbel (2007).

To explain how we deal with incoherence, we first report here the moment inequalities predicted by our model as derived in Section 4.2 of the main paper:

$$\begin{aligned}
&\mathbb{E}_{\mathrm{Pr}}\left[\Delta\Pi_{(+ab),f}(1 - G_{ab,f})Z_{(+ab),f,r} + \underline{\mathcal{L}}_r^{+} G_{ab,f} Z_{(+ab),f,r}\right] \\
&\qquad \leq \gamma_{1,f}\mathbb{E}_{\mathrm{Pr}}\left[Z_{(+ab),f,r}\right] + \gamma_{2,f}\mathbb{E}\left[\Delta\mathrm{Q}_{(+ab),f} Z_{(+ab),f,r}\right], \text{ for } r = 1, \ldots, \mathrm{R}^{+}, \\
&\mathbb{E}_{\mathrm{Pr}}\left[\Delta\Pi_{(-ab),f} G_{ab,f} Z_{(-ab),f,r} + \overline{\mathcal{U}}_r^{-}(1 - G_{ab,f})Z_{(-ab),f,r}\right] \\
&\qquad \geq \gamma_{1,f}\mathbb{E}\left[Z_{(-ab),f,r}\right] + \gamma_{2,f}\mathbb{E}_{\mathrm{Pr}}\left[\Delta\mathrm{Q}_{(-ab),f} Z_{(-ab),f,r}\right], \text{ for } r = 1, \ldots, \mathrm{R}^{-},
\end{aligned} \tag{C.1}$$

where $\mathbb{E}_{\mathrm{Pr}}$ is the expectation operator based on the probability function Pr associated with the probability space where the random variables of the model are defined. Second,

[1] A closed cycle represents a situation in which individuals never reach a stable state and constantly alternate between forming and severing links.

to simplify the exposition, we focus on *one* moment inequality from (C.1):

$$\begin{aligned}\mathbb{E}\left[\Delta\Pi_{(-ab),f}G_{ab,f}Z_{(-ab),f,r}+\overline{U}_r^{-}(1-G_{ab,f})Z_{(-ab),f,r}\right]\\ \geq \gamma_{1,f}\mathbb{E}\left[Z_{(-ab),f,r}\right]+\gamma_{2,f}\mathbb{E}\left[\Delta Q_{(-ab),f}Z_{(-ab),f,r}\right].\end{aligned} \tag{C.2}$$

Third, we streamline the notation of (C.2) as:

$$\mathbb{E}(G_mA_m)-\mathbb{E}(G_mB_m)^\top\gamma\geq 0, \tag{C.3}$$

where the subscripts f and r are omitted, m is a market $\{a,b\}$.

Let $\mathbb{P}$ be the distribution of (G_mA_m, G_mB_m) identified by the sampling process. If the set of PSNE networks is non-empty for each parameter value and variable realization, then we can replace $\mathbb{E}_{\Pr}$ with $\mathbb{E}_{\mathbb{P}}$ in (C.3) and obtain the identified set for γ associated with $\mathbb{P}$:

$$\Gamma_I \coloneqq \left\{\gamma\in\Gamma:\ \mathbb{E}_{\mathbb{P}}(G_mA_m)-\mathbb{E}_{\mathbb{P}}(G_mB_m^\top)\gamma\geq 0\right\}. \tag{C.4}$$

If the set of PSNE networks is empty for some parameter values and variable realizations, then the relationship between $\mathbb{P}$ and Pr is not completely defined because our model is silent about the realizations of (G_mA_m, G_mB_m) when the set of PSNE networks is empty. Since non-existence outcomes are never observed in our data, we approach the incoherence problem by assuming that the data are drawn from the subset of the sample space in which the set of PSNE networks is non-empty. That is, $\mathbb{P}$ comes from a truncated version of Pr, as discussed in Section 4.2 of Chesher and Rosen (2020). In what follows, we show that the identified set for γ associated with $\mathbb{P}$ is still defined by (C.4).

For ease of explanation, let us assume that A_m and B_m are discrete random variables. Given $\gamma\in\Gamma_I$, our model predicts that

$$\sum_{a\in\mathcal{A}} a\times\Pr(A_m=a,G_m=1)-\sum_{b\in\mathcal{B}}b^\top\times\Pr(B_m=b,G_m=1)\times\gamma\geq 0, \tag{C.5}$$

where $\mathcal{A}$ and $\mathcal{B}$ are the supports of A_m and B_m, respectively. Let $\mathcal{S}_{\theta,\gamma}(X^\oplus,W^\oplus,\mathrm{MS},\eta)$ be the random closed set of PSNE networks.[2] If our model is correctly specified, then the observed realization of G is associated with realizations of $X^\oplus, W^\oplus, \mathrm{MS}, \eta$ from the truncated support $\{(x^\oplus,w^\oplus,ms,\bar\eta)\in\mathrm{Supp}_{X^\oplus,W^\oplus,\mathrm{MS},\eta}:\mathcal{S}_{\theta,\gamma}(x^\oplus,w^\oplus,ms,\bar\eta)\neq\emptyset\}$. Therefore, it holds that:

$$\begin{aligned}&\mathbb{P}(A_m=a,G_m=1)=\Pr(A_m=a,G_m=1|\mathcal{S}_{\theta,\gamma}(X^\oplus,W^\oplus,\mathrm{MS},\eta)\neq\emptyset)\\ &=\frac{\Pr(A_m=a,G_m=1,\mathcal{S}_{\theta,\gamma}(X^\oplus,W^\oplus,\mathrm{MS},\eta)\neq\emptyset)}{\Pr(\mathcal{S}_{\theta,\gamma}(X^\oplus,W^\oplus,\mathrm{MS},\eta)\neq\emptyset)}=\frac{\Pr(A_m=a,G_m=1)}{\Pr(\mathcal{S}_{\theta,\gamma}(X^\oplus,W^\oplus,\mathrm{MS},\eta)\neq\emptyset)}.\end{aligned} \tag{C.6}$$

[2]For the formal definition of a random closed set, see Molchanov and Molinari (2018) and Molinari (2020).

In turn, we can write:

$$\begin{aligned}\Pr(A_m = a, G_m = 1) &= \mathbb{P}(A_m = a, G_m = 1) \times \Pr(\mathcal{S}_{\theta,\gamma}(X^{\oplus}, W^{\oplus}, \text{MS}, \eta) \neq \emptyset), \\ \Pr(B_m = b, G_m = 1) &= \mathbb{P}(B_m = b, G_m = 1) \times \Pr(\mathcal{S}_{\theta,\gamma}(X^{\oplus}, W^{\oplus}, \text{MS}, \eta) \neq \emptyset).\end{aligned} \tag{C.7}$$

We plug (C.7) in (C.5) and obtain:

$$\Pr(\mathcal{S}_{\theta,\gamma}(X^{\oplus}, W^{\oplus}, \text{MS}, \eta) \neq \emptyset) \times [\mathbb{E}_{\mathbb{P}}(G_m A_m) - \mathbb{E}_{\mathbb{P}}(G_m B_m^{\top})\gamma] \geq 0, \tag{C.8}$$

which is equivalent to

$$\mathbb{E}_{\mathbb{P}}(G_m A_m) - \mathbb{E}_{\mathbb{P}}(G_m B_m^{\top})\gamma \geq 0. \tag{C.9}$$

Hence, the identified set associated with $\mathbb{P}$ is:

$$\Gamma_I \coloneqq \Big\{\gamma \in \Gamma : \ \mathbb{E}_{\mathbb{P}}(G_m A_m) - \mathbb{E}_{\mathbb{P}}(G_m B_m^{\top})\gamma \geq 0\Big\}, \tag{C.10}$$

as in (C.4).

D Computing the First-Stage Moment Inequalities

We provide some directions on how to compute the variable $\Delta\Pi_{(+ab),f}$ in (11). A similar procedure can be followed to compute $\Delta\Pi_{(-ab),f}$ in (12). We proceed in three steps.

First, from the vector of second-stage estimates, $\hat{\theta}$, we compute the second-stage shocks for each product offered using the BLP inversion. For each airline f, we compute the mean and variance of the second-stage shocks just obtained and denote them by μ_f and Σ_f, respectively. For each potential product of each airline f, we take 100 random draws from a normal distribution with mean μ_f and variance Σ_f. We store all such draws in a matrix Ξ.

Second, we compute the expected variable profits of airline f under $(G_{(+ab),f}, G_{-f})$. To do so, we update the list of products offered by firm f, by adding direct flights between cities a and b. Further, note that setting $G_{ab,f} = 1$ induces ripple effects in neighboring markets due to the creation of new products and changes in the characteristics of existing products of airline f. For example, suppose a is a hub. Hence, adding a direct flight between cities a and b creates a one-stop flight via a between b and any city c directly connected to a, effectively adding a new product in all markets $\{b, c\}$ such that $G_{ac,f} = 1$. Additionally, the characteristics of existing products offered in neighboring markets change due to the spillover variables. In particular, we must update the demand attractiveness of these flights due to the "*Nonstop Origin*" spillover variable on the demand side. Moreover, the marginal costs of all flights originating from or destined to a or b change due to the "*Connections*" spillover variable on the marginal

cost side.

Third, let $\mathcal{M}_{ab,f}$ be the list of markets containing either new products or products with updated covariates of airline f. For each market $m \in \mathcal{M}_{ab,f}$, we let the firms re-optimize their prices by iterating on the F.O.C.s in (4), for every draw of the second-stage shocks stored in the matrix Ξ.[3] We compute the variable profits of airline f, average across draws, and get the simulated expected variable profits of airline f, which we denote by

$$\sum_{m \in \mathcal{M}_{ab,f}} \mathbb{E}[\Pi_{m,f}(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G_{(+ab),f}, G_{-f}; \theta) | X^{\oplus}, W^{\oplus}, \text{MS}].$$

We implement a similar procedure to compute the expected variable profits of airline f in each market $m \in \mathcal{M}_{ab,f}$ under G, which we denote by

$$\sum_{m \in \mathcal{M}_{ab,f}} \mathbb{E}[\Pi_{m,f}(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G; \theta) | X^{\oplus}, W^{\oplus}, \text{MS}].$$

Lastly, we calculate

$$\begin{aligned}\Delta\Pi_{(+ab),f} = &\sum_{m \in \mathcal{M}_{ab,f}} \mathbb{E}[\Pi_{m,f}(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G_{(+ab),f}, G_{-f}; \theta) | X^{\oplus}, W^{\oplus}, \text{MS}] \\ &- \sum_{m \in \mathcal{M}_{ab,f}} \mathbb{E}[\Pi_{m,f}(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G; \theta) | X^{\oplus}, W^{\oplus}, \text{MS}].\end{aligned}$$

E Bounds under Two-link Deviations

In this section, we highlight that several classes of two-link deviations do not significantly tighten our identified set. Specifically, in Section E.1, we theoretically show that two-link deviations involving either the addition or deletion of links in two non-hub markets, or in two markets sharing a hub endpoint, generate redundant inequalities compared to those produced by one-link deviations. In Section E.2, we construct the identified set for the first-stage parameters by considering two-link deviations that involve either adding a link in one hub market while deleting one in a non-hub market, or vice versa, and we find that the resulting bounds do not practically improve.

[3] We have decided to use the F.O.C.s in (4) as a contraction mapping. While we do not formally prove that (4) is indeed a contraction mapping, we have found that the resulting price vector does not change when starting from different values and that the mapping converges in all the cases considered.

E.1 Two Non-hub Markets or Two Markets Sharing a Hub

Adding Two Links

Consider markets $\{a, b\}$ and $\{c, d\}$ that are not served by airline f with direct flights (i.e., $G_{ab,f} = G_{cd,f} = 0$). From the revealed preference principle, it holds that

$$\Delta\Pi_{(+ab),f} \leq \Delta\text{FC}_{(+ab),f}, \tag{E.1}$$

$$\Delta\Pi_{(+cd),f} \leq \Delta\text{FC}_{(+cd),f}, \tag{E.2}$$

$$\Delta\Pi_{(+ab,+cd),f} \leq \Delta\text{FC}_{(+ab,+cd),f}, \tag{E.3}$$

where $\Delta\Pi_{(+ab),f}$, $\Delta\Pi_{(+cd),f}$, $\Delta\text{FC}_{(+ab),f}$, and $\Delta\text{FC}_{(+cd),f}$ are defined in Section 4.2 of the paper; $\Delta\Pi_{(+ab,+cd),f}$ is the difference between the expected variable profits at $G_{(+ab,+cd),f}$ and G_f and $\Delta\text{FC}_{(+ab,+cd),f}$ is the difference between the fixed costs at $G_{(+ab,+cd),f}$ and G_f. (E.1) and (E.2) are taken into account by our identification methodology, as they refer to one-link deviations. (E.3) is ignored by our identification methodology, as it refers to a two-link deviation. In what follows, we show that if markets $\{a, b\}$ and $\{c, d\}$ are non-hub markets for airline f and have no cities in common, or they share a hub endpoint, then (E.1) and (E.2) imply (E.3). Hence, (E.3) is redundant.

First, consider the case where markets $\{a, b\}$ and $\{c, d\}$ are non-hub markets for airline f and have no cities in common. Given our fixed cost equation, it holds that

$$\text{FC}_f(G_{(+ab,+cd),f}, \eta_f; \gamma) - \text{FC}_f(G_{(+cd),f}, \eta_f; \gamma) = \Delta\text{FC}_{(+ab),f}.$$

Therefore, the right-hand-side of (E.3) is equal to

$$\Delta\text{FC}_{(+ab,+cd),f} = \Delta\text{FC}_{(+cd),f} + \Delta\text{FC}_{(+ab),f}. \tag{E.4}$$

Observe that the left-hand-side of (E.3) can be rewritten as

$$\begin{aligned}
&\mathbb{E}[\Pi_f(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G_{(+ab,+cd),f}, G_{-f}; \theta)|X^{\oplus}, W^{\oplus}, \text{MS}] \\
&- \mathbb{E}[\Pi_f(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G_{(+cd),f}, G_{-f}; \theta)|X^{\oplus}, W^{\oplus}, \text{MS}] \\
&+ \Delta\Pi_{(+cd),f}.
\end{aligned}$$

Furthermore, from our second-stage estimates, it generally holds that

$$\begin{aligned}
&\mathbb{E}[\Pi_f(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G_{(+ab,+cd),f}, G_{-f}; \theta)|X^{\oplus}, W^{\oplus}, \text{MS}] \\
&- \mathbb{E}[\Pi_f(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G_{(+cd),f}, G_{-f}; \theta)|X^{\oplus}, W^{\oplus}, \text{MS}] \\
&\leq \Delta\Pi_{(+ab),f}.
\end{aligned} \tag{E.5}$$

In other words, adding an independent edge $\{a, b\}$ to the counterfactual network $G_{(+cd),f}$

does not tend to generate more expected variable profits than adding it to the actual network G_f. In fact, adding $\{a, b\}$ to $G_{(+cd),f}$ increases expected variable profits due to two effects. First, the demand in market $\{a, b\}$ increases because the passengers of market $\{a, b\}$ can now fly directly between a and b instead of flying through a hub of f which is neither c nor d (recall the variable "*Indirect*" entering the demand function). Second, the demand in markets having a or b as endpoints is increased by adding the direct service between a and b (recall the variable "*Nonstop Origin*" entering the demand function). From Table 5 (demand panel) we can see that the first effect dominates the second: flying direct increases utility by 1.794; adding *one* direct connection increases utility by 0.00868. In turn, by combining (E.1), (E.2) and (E.5), we see that

$$\begin{aligned}
&\mathbb{E}[\Pi_f(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G_{(+ab,+cd),f}, G_{-f}; \theta)|X^{\oplus}, W^{\oplus}, \text{MS}] \\
&- \mathbb{E}[\Pi_f(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G_{(+cd),f}, G_{-f}; \theta)|X^{\oplus}, W^{\oplus}, \text{MS}] \\
&+ \Delta\Pi_{(+cd),f} \\
&\leq \Delta\text{FC}_{(+ab),f} + \Delta\text{FC}_{(+cd),f}.
\end{aligned} \tag{E.6}$$

Hence, by combining (E.4) and (E.6), (E.3) is verified.

Second, consider the case where markets $\{a, b\}$ and $\{c, d\}$ share a hub endpoint. For instance suppose $a = c$ and a is a hub. Then,

$$\Delta\text{FC}_{(+ab,+cd),f} = \Delta\text{FC}_{(+ab),f} + \Delta\text{FC}_{(+cd),f} + \gamma_{2,f}(2N_{a,f} + 3),$$

where $N_{a,f}$ is the number of hub a's spokes in the factual network G_f. Again, given our second-stage estimates, it generally holds that

$$\begin{aligned}
&\mathbb{E}[\Pi_f(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G_{(+ab,+cd),f}, G_{-f}; \theta)|X^{\oplus}, W^{\oplus}, \text{MS}] \\
&- \mathbb{E}[\Pi_f(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G_{(+cd),f}, G_{-f}; \theta)|X^{\oplus}, W^{\oplus}, \text{MS}] \\
&- \Delta\Pi_{(+ab),f},
\end{aligned}$$

is small, compared to $\gamma_{2,f}(2N_{a,f} + 3)$. (E.5) is not always satisfied because adding $\{a, b\}$ and $\{a, d\}$ creates opportunities to fly from b to d via a. However, in our data, it is always possible to fly from b to d via other hubs in the factual network for the same airline f. As a result, it is reasonable to believe that (E.5) holds for most, if not all, two-link deviations. Therefore, using the same steps as above, we conclude that (E.3) holds.

Removing Two Links

Consider the case where markets $\{a, b\}$ and $\{c, d\}$ are served by airline f with direct flights (i.e. $G_{ab,f} = G_{cd,f} = 1$). From the revealed preference principle we can see that

$$\Delta\Pi_{(-ab),f} \geq \Delta\text{FC}_{(-ab),f}, \tag{E.7}$$

$$\Delta\Pi_{(-cd),f} \geq \Delta\text{FC}_{(-cd),f}, \tag{E.8}$$

$$\Delta\Pi_{(-ab,-cd),f} \geq \Delta\text{FC}_{(-ab,-cd),f}, \tag{E.9}$$

where $\Delta\Pi_{(-ab),f}$, $\Delta\Pi_{(-cd),f}$, $\Delta\text{FC}_{(-ab),f}$, and $\Delta\text{FC}_{(-cd),f}$ are defined in Section 4.2 of the paper; $\Delta\Pi_{(-ab,-cd),f}$ is the difference between the expected variable profits at G_f and $G_{(-ab,-cd),f}$ and $\Delta\text{FC}_{(-ab,-cd),f}$ is the difference between the fixed costs at G_f and $G_{(-ab,-cd),f}$. (E.7) and (E.8) are taken into account by our identification methodology, as they refer to one-link deviations. (E.9) is ignored by our identification methodology, as it refers to a two-link deviation. By mirroring the steps above, it is possible to show that, in most of the cases, (E.9) is redundant.

E.2 One Hub market and One Non-hub Market

Consider markets $\{a, b\}$ and $\{c, d\}$ for which $G_{ab,f} = 0$ and $G_{cd,f} = 1$. Let airline f's counterfactual network be the network in which airline f operates in all markets served under G_f, but no longer offers direct flights between cities c and d and, additionally, offers direct flights between cities a and b. We denote it by $G_{(+ab,-cd),f}$.

From the revealed preference principle, the difference between the expected variable profits earned by airline f under its counterfactual network $G_{(+ab,-cd),f}$ and the expected variable profits earned by airline f under its factual network G_f, while the competitors maintain G_{-f}, must be less than or equal to the extra fixed-cost that airline f pays for the added and subtracted direct flights:

$$\Delta\Pi_{(+ab,-cd),f} \leq \Delta\text{FC}_{(+ab,-cd),f}, \tag{E.10}$$

where $\Delta\Pi_{(+ab,-cd),f}$ is the difference between the expected variable profits at $G_{(+ab,-cd),f}$ and G_f and $\Delta\text{FC}_{(+ab,-cd),f}$ is the difference between the fixed costs at $G_{(+ab,-cd),f}$ and G_f. (E.10) is not taken into account by our identification methodology, as it refers to a two-link deviation. In what follows, we construct the estimated identified set for the first-stage parameters based on (E.10), assuming that $\{a, b\}$ is a hub market and $\{c, d\}$ is a non-hub market, or vice versa.

Before proceeding, we make two remarks. First, note that (E.10) could, in principle, generate useful restrictions on the congestion cost parameter, $\gamma_{2,f}$, given our fixed cost

specification. For example, when a is a hub and b, c, and d are non-hubs, then

$$\Delta \text{FC}_{(+ab,-cd),f} = \eta_{ab,f} - \eta_{cd,f} + \gamma_{2,f}(2N_{a,f} + 1).$$

Hence, (E.10) can provide a "lower" bound for $\gamma_{2,f}$. Similarly, when c is a hub and a, b, and d are non-hubs, then

$$\Delta \text{FC}_{(+ab,-cd),f} = \eta_{ab,f} - \eta_{cd,f} + \gamma_{2,f}(1 - 2N_{c,f}),$$

Hence, (E.10) can provide an "upper" bound for $\gamma_{2,f}$.

Second, as thoroughly discussed in Section 4.2 of the paper, in order to transform (E.10) into a moment inequality, we rely on the availability of one (or more) exogenous binary "instruments", denoted $Z_{(+ab,-cd),f}$, which may vary by both market and firm and satisfy a standard exogeneity condition:

$$\mathbb{E}\big(\eta_{ab,f}, \eta_{cd,f} \mid Z_{(+ab,-cd),f} = 1\big) = 0.$$

To construct such instruments, we combine the covariates $Z_{(+ab),f}$ and $Z_{(-cd),f}$ defined in Section 4.2 and used for one-link deviations. In particular, we define $Z_{(+ab,-cd),f}$ to take the value one if both $Z_{(+ab),f} = 1$ and $Z_{(-cd),f} = 1$, and zero otherwise. In doing so, as indicated in Table 3, we construct 7 instruments per airline.

Table E.1 displays the projection for $\gamma_{2,f}$ of the estimated identified set of the first-stage parameters, $\widehat{\Gamma}_I$ (defined in (25)), constructed based on the moment inequalities derived from (E.10). Comparing these results with those in Table 7, we observe that the projections are not refined by the two-link deviations considered.

Table E.1: First-stage estimates of $\gamma_{2,f}$ based on two-link deviations.

Congestion costs ($\gamma_{2,f}$)	**LB**	**UB**
AA	3,018	36,294
DL	1,326	29,048
UA	1,554	22,106
US	1,556	35,090
WN	1,342	39,988

Note: Quantities are in USD.

F Inference on the Demand and Supply Parameters

We conduct inference on θ via GMM under the assumption that the number of markets goes to infinity. Formally, we consider the moment conditions of Section 4.1 and use their sample analogues to construct a GMM objective function which should be minimized with

respect to $\theta \in \Theta$:

$$Q(\theta) = D(\theta)' A D(\theta), \tag{F.1}$$

where

$$D(\theta) \coloneqq \begin{pmatrix} \frac{1}{|\mathcal{J}|} \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{J}_m} [\tau_{j,m}(X_m^{\oplus}, W_m^{\oplus}, \mathrm{MS}_m, s_m^{\oplus}, P_m^{\oplus}, G; \theta) \times z_{j,m,1}(X_m^{\oplus}, W_m^{\oplus})] \\ \frac{1}{|\mathcal{J}|} \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{J}_m} [\tau_{j,m}(X_m^{\oplus}, W_m^{\oplus}, \mathrm{MS}_m, s_m^{\oplus}, P_m^{\oplus}, G; \theta) \times z_{j,m,2}(X_m^{\oplus}, W_m^{\oplus})] \\ \vdots \\ \frac{1}{|\mathcal{J}|} \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{J}_m} [\tau_{j,m}(X_m^{\oplus}, W_m^{\oplus}, \mathrm{MS}_m, s_m^{\oplus}, P_m^{\oplus}, G; \theta) \times z_{j,m,\mathrm{L}}(X_m^{\oplus}, W_m^{\oplus})] \end{pmatrix},$$

$\mathcal{J} \coloneqq \cup_{m \in \mathcal{M}} \mathcal{J}_m$ is the set of all offered products, and A is an appropriate $2\mathrm{L} \times 2\mathrm{L}$ weighting matrix. In particular, A is computed via the usual two-step procedure: first, we estimate the parameters using the optimal weighting matrix under conditional homoskedasticity; second, we use the obtained estimates to construct the optimal weighting matrix under conditional heteroskedasticity and re-estimate the parameters.

Note that we estimate the demand and supply sides jointly. We could also estimate the demand and supply sides separately by following a two-step procedure: first estimating the demand parameters; then using these estimates to calculate the markups; finally regressing the resulting marginal costs on the observed marginal cost shifts to obtain the supply parameters. We have chosen to estimate the demand and supply sides together because it allows us to take into account the potential correlation between the demand and supply moments and thus obtain more precise estimates, as discussed in Berry et al. (1995). Moreover, since we have a computationally "light" demand specification, the additional cost of estimating the demand and supply sides jointly is negligible.

G Spatial HAC Estimation of the Variance

In Sections G.1 and G.2, we explain how to construct confidence intervals for both the first- and second-stage parameters without assuming that data points are mutually independent across markets. In fact, in our setting, the observed product characteristics offered by a given airline in different markets are correlated due to spillover variables, which in turn induce correlation among the moment functions used for identification. To address this issue, we employ a spatial heteroscedasticity and autocorrelation consistent (SHAC) estimator for the variances.

Constructing a HAC estimator of the variance generally requires defining a distance measure between units of observation. In the network literature, the typical unit is a node (a city, in our context), so HAC estimators are often based on a distance measure between nodes; see, for example, Kojevnikov (2021), Kojevnikov et al. (2021), and Leung (2023). By contrast, our framework considers a market (a pair of nodes) as the unit of observation, necessitating a distance measure between markets. The network literature

commonly uses *path distance*, defined as the length of the shortest path connecting two nodes a and b. However, this notion does not readily extend to pairs of nodes. For instance, we attempted to adapt path distance to markets $\{a, b\}$ and $\{c, d\}$ by taking the maximum of the length of the shortest paths among their endpoints, but these measures were nearly identical across most pairs of markets. Consequently, they did not effectively reflect a decaying distance measure in our setting. As a result, we chose a *geographical* distance measure (Conley, 2010), since it is natural in our empirical context to assume that the correlation between markets systematically declines with geographic distance.

Lastly, in Section G.3, we explain how to incorporate the uncertainty coming from the second-stage estimates when constructing confidence regions for the first-stage parameters.

G.1 Demand and Supply Parameters

Consider the moment equalities listed in (10) in Section 4.1 of the paper:

$$\mathbb{E}\big(\tau_{j,m}(\theta) \times z_{j,m,l}\big) = 0 \quad \forall\, l = 1, \ldots, \mathrm{L},$$

where we have omitted the arguments $(X_m^{\oplus}, W_m^{\oplus}, \mathrm{MS}_m, s_m^{\oplus}, P_m^{\oplus}, G)$ to streamline the notation. Let

$$g_{j,m,l}(\theta) \coloneqq \tau_{j,m}(\theta) \times z_{j,m,l},$$

and

$$\overline{g}_l(\theta) \coloneqq \frac{1}{|\mathcal{J}|} \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{J}_m} g_{j,m,l}(\theta).$$

In what follows, we describe how to estimate, for each $\theta \in \Theta$, the variance-covariance matrix of the vector

$$\overline{g}(\theta) \coloneqq (\overline{g}_l(\theta) : l = 1, \ldots, \mathrm{L}).$$

This variance–covariance matrix has size $2\mathrm{L} \times 2\mathrm{L}$, where 2L is the total number of instrument from both the demand and supply sides. The estimation of this variance-covariance matrix requires consideration of the fact that the data are not mutually independent across markets. To address this correlation, we employ a spatial heteroskedasticity and autocorrelation consistent (SHAC) estimator.

In particular, an estimator of the covariance between $\overline{g}_l(\theta)$ and $\overline{g}_{l'}(\theta)$ for any two $l, l' = 1, \ldots, \mathrm{L}$ is:

$$\widehat{\mathrm{Cov}}(\overline{g}_l(\theta), \overline{g}_{l'}(\theta)) = \frac{1}{|\mathcal{J}|^2} \sum_{(m,m') \in \mathcal{M}^2} \sum_{j \in \mathcal{J}_m} \sum_{j' \in \mathcal{J}_{m'}} k(j, j')\, g_{j,m,l}(\theta) g^{\top}_{j',m',l'}(\theta) - \overline{g}_l(\theta)\overline{g}_{l'}(\theta) \quad \text{(G.1)}$$

where $k(j, j')$ is a kernel function that depends on an appropriately defined distance (see

below) between the two markets m and m', and is nonzero if the products j, j' are offered by the same airline.[4]

We calculate the distance between two markets as follows. We project the United States onto a plane using a Mercator projection that is tangent to the Earth at the center of the U.S. Representing a market as a segment $[a, b]$, the distance between two markets, $[a, b]$ and $[c, d]$, is measured using the Hausdorff distance from convex theory:

$$d_H([a,b],[c,d]) = \max\left\{ \max_{x\in[a,b]} d\big(x,[c,d]\big),\ \max_{y\in[c,d]} d\big(y,[a,b]\big) \right\}.$$

Finally, we employ a triangular kernel with a cutoff, $D_{\max}$, of 200 miles. Specifically, if airline f offers product j in market m and product j' in market m', then

$$k(j,j') = \left(1 - \frac{d(m,m')}{200}\right) \mathbf{1}\{d(m,m') \leq 200\},$$

and $k(j, j') = 0$ otherwise.

Table G.1 shows the estimated standard deviation for the main demand parameters, both without (third column) and with (fourth column) the SHAC correction.

Table G.1: Demand estimates.

Utility function			
	Coefficient	Std. deviation	
		i.i.d.	SHAC $D_{\max} = 200$
Intercept	-5.598	(0.235)	(0.309)
Price	-0.587	(0.066)	(0.085)
Indirect	-1.794	(0.065)	(0.088)
Nonstop Origin	0.868	(0.077)	(0.088)
Distance	0.289	(0.090)	(0.124)
Distance2	-0.093	(0.018)	(0.029)
Nesting Parameter (λ)	0.623	(0.025)	(0.035)

G.2 Fixed Cost Parameters

Consider the moment inequalities listed in (23) in Section 5 of the paper:

$$\mathbb{E}(Z_{m,r}B_m)^\top \gamma - \mathbb{E}(Z_{m,r}A_m) \leq 0, \quad r = 1, \ldots, \mathrm{R}.$$

Let

$$\tau_{m,r}(\gamma) \coloneqq Z_{m,r}B_m^\top \gamma - Z_{m,r}A_m,$$

[4]See Conley (2010).

and

$$\overline{\tau_r}(\gamma) \coloneqq \frac{1}{\mathrm{M}} \sum_{m=1}^{\mathrm{M}} \Big(Z_{m,r} B_m^\top \gamma - Z_{m,r} A_m \Big).$$

We now describe how to estimate, for each $\gamma \in \Gamma$, the variance-covariance matrix of the vector

$$\overline{\tau}(\gamma) \coloneqq (\overline{\tau_r}(\gamma) : r = 1, \ldots, \mathrm{R}).$$

This variance-covariance matrix has size 28×28, where 28 is the number of moment inequalities used in our estimation procedure. Once again, the estimation of this variance-covariance matrix requires accounting for the fact that the data are not mutually independent across markets. To address this correlation, we use the SHAC estimator presented in Section G.1.

In particular, following (G.1), an estimator of the covariance between $\overline{\tau_r}(\gamma)$ and $\overline{\tau_{r'}}(\gamma)$ for any two $r, r' = 1, \ldots, \mathrm{R}$ is:

$$\widehat{\mathrm{Cov}}(\overline{\tau_r}(\gamma), \overline{\tau_{r'}}(\gamma)) = \frac{1}{\mathrm{M}^2} \sum_{(m,m') \in \mathcal{M}^2} k(m,m') \tau_{m,r}(\gamma) \tau_{m',r'}(\gamma) - \overline{\tau_r}(\gamma) \overline{\tau_{r'}}(\gamma),$$

where

$$k(m,m') = \left(1 - \frac{d(m,m')}{200}\right) \mathbf{1}\{d(m,m') \leq 200\}.$$

With the variance–covariance matrix of $\overline{\tau}(\gamma)$, $\Sigma(\gamma)$, computed, we can subsequently estimate, for any direction q, the variance $v_\kappa(q)$ of $\mathbb{G}_\kappa(q)$. Recall that $\mathbb{G}_\kappa(q)$ is the limit in distribution of $\hat{\delta}(q; \Gamma_I^\kappa)$ as stated in Theorem 2, and that $v_\kappa(q)$ is necessary for computing both the confidence interval for each component of γ and the confidence region for the entire vector.

Specifically, for a given direction q, let γ_q^κ denote the unique element of Γ_I^κ that lies on the supporting hyperplane in direction q, and let λ_q^κ be the Lagrange multiplier obtained when estimating the support function in direction q. Let $s_\kappa(q)$ be the $\mathrm{R} \times 1$ vector collecting the (estimated) weights for each moment inequality:

$$s_\kappa(q) = \left[\frac{1}{1 + \exp\left(-\kappa \left[\hat{b}_1^\top \gamma_q^\kappa - \hat{a}_1\right]\right)}, \ldots, \frac{1}{1 + \exp\left(-\kappa \left[\hat{b}_\mathrm{R}^\top \gamma_q^\kappa - \hat{a}_\mathrm{R}\right]\right)} \right]^\top,$$

where

$$\hat{b}_r = \frac{1}{\mathrm{M}} \sum_{m=1}^{\mathrm{M}} Z_{m,r} B_m, \quad \hat{a}_r = \frac{1}{\mathrm{M}} \sum_{m=1}^{\mathrm{M}} Z_{m,r} A_m, \quad r = 1, \ldots, \mathrm{R}.$$

Following Theorem 2, $v_\kappa(q)$ can be estimated as

$$\widehat{v_\kappa(q)} = (\lambda_q^\kappa)^2 \, s_\kappa(q)^\top \Sigma(\gamma_q^\kappa) s_\kappa(q). \tag{G.2}$$

G.3 Uncertainty from the Second Stage

When estimating the variance-covariance matrix of $\overline{\tau}(\gamma)$ in Section G.2, note that A_m is computed using the quantity $\Delta\Pi_{(+ab),f}$ (or $\Delta\Pi_{(-ab),f}$), as discussed in Section D. Since this quantity depends on the second-stage parameters θ (of which we only have an estimate, $\hat{\theta}$), we must account for this additional source of uncertainty in our variance estimation.

To do so, we draw values from the 95% confidence region for θ and denote them by $\{\hat{\theta}_i\}_i$.[5] For each $\hat{\theta}_i$, we follow the procedure described in Section D to obtain an estimate of $\Delta\Pi_{(+ab),f}$, which we denote by $\Delta\Pi_{(+ab),f}(\hat{\theta}_i)$ (or similarly for $\Delta\Pi_{(-ab),f}$). Using the decomposition

$$\Delta\Pi_{(+ab),f}(\hat{\theta}) = \Delta\Pi_{(+ab),f}(\theta^0) + \text{noise}_{(+ab),f},$$

where θ^0 denotes the true unkown value of θ, the collection $\{\Delta\Pi_{(+ab),f}(\hat{\theta}_i)\}_i$ enables us to estimate the noise term, $\widehat{\text{noise}}_{(+ab),f}$. Finally, we compute the SHAC estimator of the variance-covariance matrix of $\overline{\tau}(\gamma)$ by adding the variance of $\widehat{\text{noise}}_{(+ab),f}$ to the calculations described in Section G.2.

Table G.2 illustrates the effect of accounting for the noise in the second-stage regression as opposed to ignoring it. Because the demand and supply estimates are relatively precise, the 95% confidence region of the first-stage parameters is largely determined by the first-stage inference procedure.

Table G.2: First-stage inference with and without uncertainty from second stage estimates.

	95% Conf . Interval			
	With 2nd stage uncertainty		Without 2nd stage uncertainty	
	LB	UB	LB	UB
Baseline fixed costs ($\gamma_1.f$)				
Legacy carriers (AA, US, DL, UA)	554,483	1,299,833	554,780	1,295,108
WN	808,732	2,173,262	809,420	2,171,478
Congestion costs ($\gamma_{2,f}$)				
AA	5,150	41,203	5,256	41,159
DL	4,158	31,740	4,241	31,672
UA	3,542	19,780	3,609	19,705
US	11,199	39,159	11,277	39,097
WN	14,339	39,804	14,502	39,422

Note: Quantities are in USD.

[5] Recall that the second-stage estimator is a GMM estimator and is therefore regular; hence its 95% confidence region can be derived from standard GMM theory.

H Inference on the Fixed cost Parameters

H.1 Writing (28) as Linear Optimization Problem with Exponential Cone Constraints

In what follows, we show that (28) is a linear optimization problem with exponential cone constraints. First, we simplify the notation of (28) and write it as

$$\begin{aligned} \delta(q, \Gamma_I^\kappa) \coloneqq \sup_{\gamma \in \Gamma} & \, q^\top \gamma, \\ \text{s.t.} & \sum_{r=1}^{\mathrm{R}} f_\kappa(b_r \gamma - a_r) - \mathrm{R} \log(2)/\kappa \leq 0, \end{aligned} \tag{H.1}$$

where b_r stands for $\mathbb{E}(Z_{m,r} B_m)^\top$ and a_r for $\mathbb{E}(Z_{m,r} A_m)$. Both quantities can be estimated consistently from their empirical analogue. Second, observe that

$$\sum_{r=1}^{\mathrm{R}} f_\kappa(b_r \gamma - a_r) - \mathrm{R} \log(2)/\kappa \leq 0 \tag{H.2}$$

$$\Leftrightarrow \log(1 + \exp(\kappa(b_r \gamma - a_r))) \leq t_r \text{ for } r = 1, \ldots, \mathrm{R} \text{ and } \sum_{r=1}^{\mathrm{R}} t_r \leq \mathrm{R} \log 2 \tag{H.3}$$

$$\Leftrightarrow \exp(-t_r) + \exp(-t_r + \kappa(b_r \gamma - a_r)) \leq 1 \text{ for } r = 1, \ldots, \mathrm{R} \text{ and } \sum_{r=1}^{\mathrm{R}} t_r \leq \mathrm{R} \log 2. \tag{H.4}$$

Therefore, (H.1) is equivalent to

$$\max q^\top \gamma + \sum_{r=1}^{R} 0 \times t_r + 0 \times u_r + 0 \times v_r,$$

under the constraints

$$\begin{aligned} \sum_{r=1}^{\mathrm{R}} t_r &\leq \mathrm{R} \log 2, \\ u_r + v_r &\leq 1, \; r = 1, \ldots, \mathrm{R}, \\ (v_r, 1, -t_r) &\in K_{\exp} \; r = 1, \ldots, \mathrm{R}, \\ (u_r, 1, \kappa(b_r \gamma - a_r) - t_r) &\in K_{\exp} \; r = 1, \ldots, \mathrm{R}. \end{aligned}$$

The exponential cone $K_{\exp}$ is a convex subset of $\mathbb{R}^3$ such that

$$K_{\exp} = \{(x_1, x_2, x_3) : \; x_1 \geq x_2 \exp(x_3/x_2); x_2 > 0\} \cup \{(x_1, 0, x_3), \; x_1 \geq 0, x_3 \geq 0\}.$$

The constraints above ensure, in particular, that for any r, $v_r \geq \exp(-t_r)$ and $u_r \geq \exp(-t_r + \kappa(b_r\gamma - a_r))$, and, therefore, ensure (H.4).

See `https://docs.mosek.com/modeling-cookbook/expo.html#softplus-function` for further details.

H.2 Discussion of the Smoothing Strategy

In this section, we discuss the advantages of smoothing the identified set by transforming the system of R linear inequality constraints (23) into a unique inequality constraint (27), as discussed in Section 5 of the paper.

Theorem 1 shows that the estimated support function in a given direction q does not always converge to a standard normal random variable. This irregularity motivates the use of smoothing techniques.

To build intuition for this result, consider the case where we estimate the support function in direction q for a two-dimensional polytope Γ_I defined by R inequalities:

$$b_r^\top \gamma - a_r \leq 0, \quad r = 1, \ldots, \mathrm{R}.$$

Each inequality is estimated from a sample of M observations, with:

$$\hat{b}_r = \frac{1}{\mathrm{M}} \sum_{m=1}^{\mathrm{M}} b_{m,r}, \quad \hat{a}_r = \frac{1}{\mathrm{M}} \sum_{m=1}^{\mathrm{M}} a_{m,r}.$$

Following the notation of our original moment inequalities in (23), note that we abbreviate $Z_{m,r}B_m$ as $b_{m,r}$ and $Z_{m,r}A_m$ as $a_{m,r}$. Similarly, we denote $\mathbb{E}(Z_{m,r}B_m)$ by b_r and $\mathbb{E}(Z_{m,r}A_m)$ by a_r.

Assume that all R inequalities are active in defining the identified set Γ_I, and that any vertex of Γ_I is formed by the intersection of exactly two inequalities (as in Figure H.1 with R = 5). Under this assumption, the set of Lagrange multipliers, $\mathcal{L}_0(q)$, defined in Section 5, is a singleton. This condition is known in the optimization literature as the Linear Independence Constraint Qualification (LICQ).

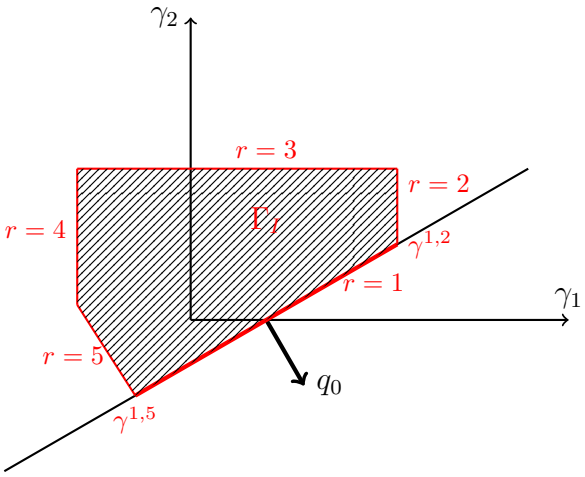


Figure H.1: A polytope Γ_I defined by R = 5 inequalities.

Consider estimating the support function in direction q_0 as depicted in Figure H.1. The intersection between the supporting hyperplane in direction q_0 and the polytope Γ_I is the segment $[\gamma^{1,2}, \gamma^{1,5}]$, which is the exposed face of the convex set in direction q_0.

According to Theorem 1, the asymptotic distribution of the estimated support function—when the five inequalities are estimated from the data—is given by:

$$\sqrt{\mathrm{M}}\left(\hat{\delta}(q_0;\Gamma_I) - \delta(q_0;\Gamma_I)\right) \xrightarrow[\mathrm{M}\to\infty]{d} \sup_{\gamma\in[\gamma^{1,2},\gamma^{1,5}]} \lambda_1 W_1(\gamma),$$

where λ_1 is the Lagrange multiplier associated with the first inequality (i.e., $r = 1$), which is the only binding inequality among the five, and $W_1(\gamma)$ is the limiting distribution of

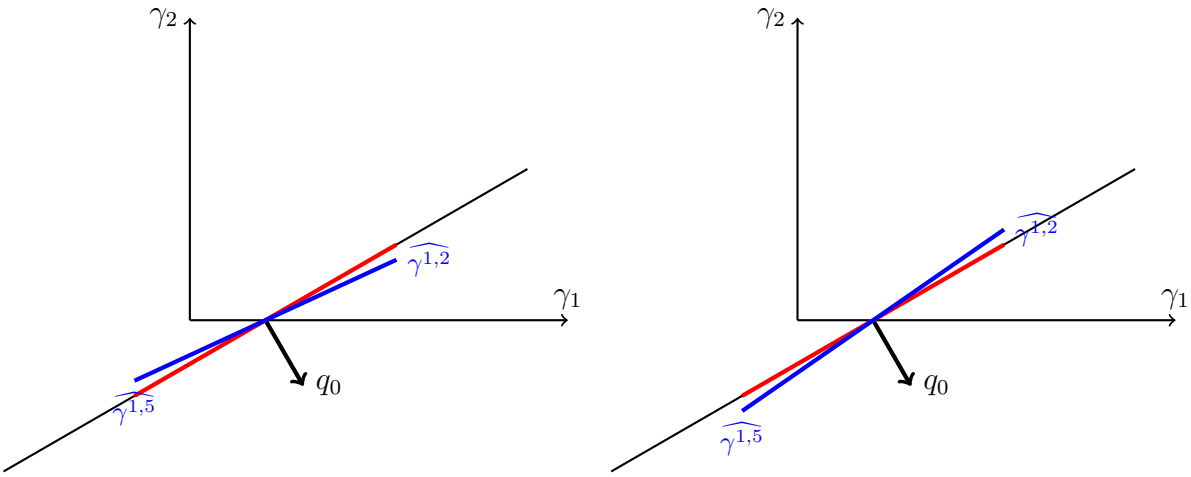
$$\frac{1}{\mathrm{M}}\sum_{m=1}^{\mathrm{M}}\left(b_{m,1}^{\top}\gamma - a_{m,1}\right),$$

a Gaussian process with variance that can be estimated using the SHAC estimator described in Section G.2. In general, this asymptotic distribution is non-normal and takes the form:

$$\max\left(\lambda_1 W_1(\gamma^{1,2}),\ \lambda_1 W_1(\gamma^{1,5})\right).$$

The reason is that, due to sampling variability, the segment $[\gamma^{1,2}, \gamma^{1,5}]$ is estimated as $[\widehat{\gamma}^{1,2}, \widehat{\gamma}^{1,5}]$ (see Figure H.2). Sometimes, the maximum of $q_0^{\top}\gamma$ is attained at $\widehat{\gamma}^{1,2}$, as illustrated in Figure H.2(a), while in other samples, it is attained at $\widehat{\gamma}^{1,5}$ as illustrated in Figure H.2(b). This variation leads to the non-normality of the asymptotic distribution.

If we were able to target a specific point on the segment $[\gamma^{1,2}, \gamma^{1,5}]$ and use this point to estimate the support function in direction q_0, we could restore normality. Take, for example, the midpoint γ_c of $[\gamma^{1,2}, \gamma^{1,5}]$ and consider the estimated midpoint as $\widehat{\gamma_c} = \frac{1}{2}(\widehat{\gamma^{1,2}} + \widehat{\gamma^{1,5}})$. The support function of Γ_I in direction q_0 is also equal to $q_0^{\top}\gamma_c$ and can



(a) $\hat{\delta}(q_0;\Gamma_I) = q_0^\top \widehat{\gamma^{1,2}}$ in the smooth case

(b) $\hat{\delta}(q_0;\Gamma_I) = q_0^\top \widehat{\gamma^{1,5}}$

Figure H.2: Exposed face and asymptotic distribution of $\hat{\delta}(q_0;\Gamma_I)$.

be estimated by $q_0^\top \widehat{\gamma}_c$. The vertices of the estimated polytope are estimators of the true vertices, and their asymptotic distribution is normal (because each estimated coefficient of each inequality is asymptotically normally distributed). Therefore, this estimator of the support function is asymptotically normal as well.

Based on this intuition, the smoothing approach implicitly selects a point on the segment $[\gamma^{1,2}, \gamma^{1,5}]$; the intersection between the hyperplane with outer normal q_0 and the smooth outer set converges to this point as $\kappa \to \infty$.

A Graphical Illustration

Consider the polytope, $\Gamma_I \in \mathbb{R}^2$ defined by the following 6 inequalities:

$$\begin{aligned}
\text{Ineq 1:} \quad & \gamma_1 \leq 2, \\
\text{Ineq 2:} \quad & \gamma_1 + \gamma_2 \leq 4, \\
\text{Ineq 3:} \quad & \gamma_2 \leq 4, \\
\text{Ineq 4:} \quad & -\gamma_1 + \gamma_2 \leq 4, \\
\text{Ineq 5:} \quad & -2\gamma_1 - \gamma_2 \leq -1, \\
\text{Ineq 6:} \quad & 2\gamma_1 - \gamma_2 \leq 4.
\end{aligned}$$

The polytope is drawn in Figure H.3, where its exposed faces are represented by the red dashed lines. We also depict the smooth outer set $\hat{\Gamma}_I^\kappa$ for five values of the smoothing parameter κ, namely $\kappa = 1, 3, 5, 10$, and 20. First, observe that the outer set tends toward the true polytope as κ increases. For each value of κ, we report the point that achieves the maximum of γ_1 over the corresponding smooth outer set. Standard calculations show that:

$$\max \gamma_1 \simeq 2 + \frac{\log(2^R - 1)}{\kappa}$$

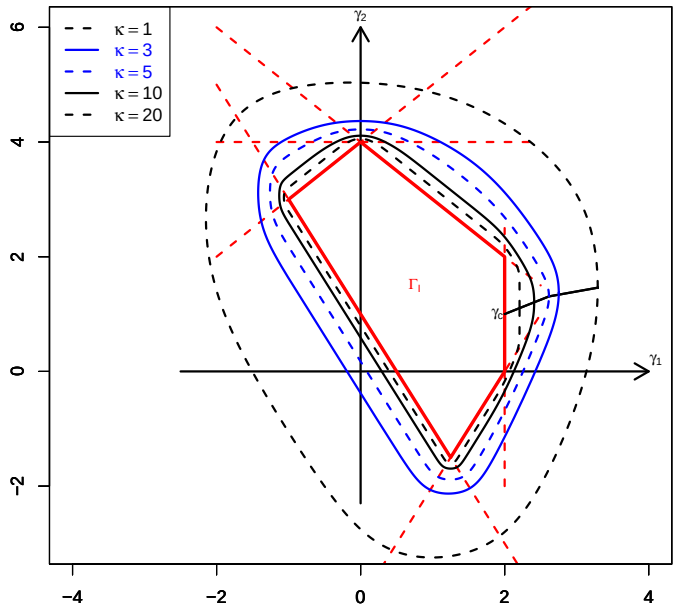


Figure H.3: Polytope Γ_I and smooth outer sets Γ_I^κ.

and $\gamma_2 \to 1$, which corresponds to the value that equalizes the constraints $\gamma_1 + \gamma_2 - 4$ and $2\gamma_1 - \gamma_2 - 4$ (the two adjacent inequalities of our exposed face) when $\gamma_1 = 2$. We also trace the theoretical points that achieve the maximum of γ_1 for values of κ ranging from 1 to 5000. This is the black curve on the right of the figure. The limiting point, denoted γ_c, is indicated in the figure. Note that the smoothing procedure implicitly selects a point on the exposed face, and this selection converges to γ_c as $\kappa \to \infty$.

Interestingly, the third inequality is not necessary to define Γ_I, and three inequalities intersect at the point $(4, 0)$. This configuration constitutes a violation of the classical LICQ (Linear Independence Constraint Qualification) assumption mentioned above. As a consequence, the estimation of the support function in the direction $(0, 1)$ is no longer asymptotically normal because the Lagrange multipliers of the corresponding optimization procedure are no longer unique. However, the smoothing approach implicitly stretches these three inequalities, thereby resolving the issue. By replacing the sharp intersection with a smooth approximation, the smoothing procedure restores regularity and avoids the degeneracy caused by the violation of LICQ.

Support Function and Asymptotic Normality in Specific Directions

Note that there is one case in which the support function in direction q_0 is asymptotically normal, even if Γ_I has an exposed face in that direction. This occurs when q_0 consistently

remains the outer normal of an exposed face of the estimated polytope for *every* random sample. We illustrate this with a specific example.

In our setting, the binding inequalities that determine the maximum and minimum values of $\gamma_{1,f}$ are known. These inequalities stem from one-link deviations in non-hub markets, where $\gamma_{2,f}$ is absent both in the sample and in the population. For instance, consider the inequality that determines the minimum value of $\gamma_{1,f}$. This inequality arises by adding a nonexistent direct flight in a non-hub market $\{a, b\}$ to airline f's network, connecting a medium-sized city to a competitor's hub. Specifically, it corresponds to the instrument $Z_{(+ab),f,3}$ defined in Table 3:

$$\gamma_{1,f}\mathbb{E}\left[Z_{(+ab),f,3}\right] \geq \mathbb{E}\left[\Delta\Pi_{(+ab),f}Z_{(+ab),f,3}\right],$$

following the notation of (22) with $Z_{(+ab),f,r} = Z_{(+ab),f,3}$, $G_{ab,f} = 0$, and without $\gamma_{2,f}$.

Denote the sample analogue of this inequality as:

$$\gamma_{1,f}\overline{Z_{(+ab),f,3}} \geq \overline{\Delta\Pi_{(+ab),f}Z_{(+ab),f,3}},$$

Assuming the limit of the sample average $\overline{Z_{(+ab),f,3}}$ is non-zero, which holds in this case, we obtain:

$$\gamma_{1,f} \geq \frac{\overline{\Delta\Pi_{(+ab),f}Z_{(+ab),f,3}}}{\overline{Z_{(+ab),f,3}}}.$$

Therefore, the estimated support function in direction $q_0 = (-1, 0, \ldots, 0)^\top$ is given by

$$-\frac{\overline{\Delta\Pi_{(+ab),f}Z_{(+ab),f,3}}}{\overline{Z_{(+ab),f,3}}},$$

which is asymptotically normal by the delta method, as both the numerator and denominator are sample averages.

We can repeat analogous arguments for the maximum value of $\gamma_{1,f}$, which corresponds to the direction $q_0 = (1, 0, \ldots, 0)^\top$ thereby obtaining that the asymptotic distribution of the support function in direction q_0 is also asymptotically normal.

Note that both estimated support functions are asymptotically normal, despite the presence of exposed faces in the directions q_0, because the estimated identified set exhibits exposed faces exactly in these directions for every random sample, rather than only approximately.

Importantly, the derivations above suggest that, in the case of $\gamma_{1,f}$ (and not $\gamma_{2,f}$), we can construct a confidence interval without needing to implement the smoothing procedure. We can leverage this special feature of our framework to further assess the performance of our inference approach. Specifically, based on the asymptotic normality, the confidence interval for $\gamma_{1,f}$ is $(0.557, 1.263)$ for a legacy carrier and $(0.816, 2.151)$ for

Southwest Airlines (in millions of dollars). Table 9 shows that the confidence intervals obtained using our smoothing procedure closely approximate these "correct" intervals.

Calibrating the Value of κ

Finally, we discuss the trade-off involved in choosing the smoothing parameter κ. A small value of κ produces a smooth outer set with slowly varying curvature (see Figure H.3); however, the distance between this smooth outer set and the true identified set—our object of interest—can be large, leading to wide confidence regions. As κ increases, this distance decreases, as demonstrated in Chen and Mangasarian (1995). Nonetheless, while the variance of the support function estimator decreases smoothly with κ, the validity of the asymptotic distribution becomes increasingly questionable for large κ.

To better understand this phenomenon, consider the second-order expansion of the inequality constraint $g_\kappa(\gamma)$ defined in (27) as $\mathrm{M} \to \infty$ for a fixed κ. Following the derivation of Theorem 2, we can write:

$$\begin{aligned}\sqrt{\mathrm{M}}\left(\hat{g}_\kappa(\gamma) - g_\kappa(\gamma)\right) =& \sqrt{\mathrm{M}} \sum_{r=1}^{\mathrm{R}} \frac{\Delta_r(\gamma)}{1 + \exp\left(-\kappa\left[b_r^\top \gamma - a_r\right]\right)} \\ &+ \frac{\kappa}{2\sqrt{\mathrm{M}}} \sum_{r=1}^{\mathrm{R}} \frac{\exp\left(-\kappa\left[b_r^\top \gamma - a_r\right]\right)}{\left(1 + \exp\left(-\kappa\left[b_r^\top \gamma - a_r\right]\right)\right)^2} W_r(\gamma)^2 + o_P(1/\sqrt{\mathrm{M}}),\end{aligned} \tag{H.5}$$

where $\Delta_r(\gamma) = (\hat{b}_r - b_r)^\top \gamma - (\hat{a}_r - a_r)$, $W_r(\gamma)$ denotes the limiting in distribution of $\sqrt{M}\Delta_r(\gamma)$—a normal distribution—and the variables $b_r, \hat{b}_r, a_r, \hat{a}_r$ are defined at the beginning of Section H.2.

The expansion in (H.5) reveals a tension between the bias (linked to the distance between the smooth and the true identified set) and the remainder term, which is of order $\kappa/\sqrt{\mathrm{M}}$.

Consider now the support function in direction q and assume q is the outer normal of an exposed face. In this case, only one inequality, say r_0, is binding. Let γ_q^κ denote the unique point of intersection between the outer set Γ_I^κ and its supporting hyperplane in direction q.

For large κ, it holds that

$$g_\kappa(\gamma_q^\kappa) \simeq \frac{1}{\kappa} \log\left(1 + \exp\left(\kappa\left[b_{r_0}^\top \gamma_q^\kappa - a_{r_0}\right]\right)\right) - \frac{\mathrm{R}\log 2}{\kappa},$$

implying

$$b_{r_0}^\top \gamma_q^\kappa - a_{r_0} \simeq \frac{\log(2^{\mathrm{R}} - 1)}{\kappa}.$$

Plugging the latter expression into the second-order expansion, the leading remainder

term becomes approximately:

$$\frac{\kappa}{\sqrt{\mathrm{M}}}\frac{2^{\mathrm{R}}-1}{(2^{\mathrm{R}})^2}W_{r_0}(\gamma)^2.$$

Using the average value of the variances of $W_r(\gamma)$ and setting $\mathrm{R} = 28$—the number of moment inequalities in our empirical application—we obtain an estimated remainder term that is negligible for values of κ below 10,000; specifically, its magnitude is less than 10^{-6} times that of the first-order term. Recall that we chose $\kappa = 100$ in our empirical application to control the distance between the outer set and the identified set. At this value, the negligible magnitude of the remainder term ensures the robustness and validity of our approximation.

H.3 Constructing a Confidence Interval for a Component of γ

Suppose we want to construct a confidence interval for a specific linear combination of components of γ, $c^\top\gamma$. Let $q = c/\|c\|$. By Theorem 2,

$$\sqrt{\mathrm{M}}\left(\hat{\delta}(q;\Gamma_I^\kappa) - \delta(q;\Gamma_I^\kappa)\right) \xrightarrow[\mathrm{M}\to\infty]{d} \mathbb{G}_\kappa(q).$$

The optimization routine detailed in Section H.1 gives us the unique point, γ_q, which achieves the maximum of $c^\top\gamma$ on $\widehat{\Gamma}_I^\kappa$. Equation (G.2) shows how to estimate the variance of $\mathbb{G}_\kappa(q)$, which is denoted by $v_\kappa(q)$. The quantity

$$c^\top\gamma_q + \|c\|n_\alpha\sqrt{v_\kappa(q)},$$

is the upper bound of the $1-\alpha$ confidence interval for $c^\top\gamma$, where n_α is the α-th quantile of the standard normal distribution.

Similarly, let $-q = -c/\|c\|$ and γ_{-q} be the point which achieves the maximum of $-c^\top\gamma$ on $\widehat{\Gamma}_I^\kappa$. The quantity

$$c^\top\gamma_{-q} - \|c\|n_\alpha\sqrt{v_\kappa(-q)},$$

is the lower bound of the $1-\alpha$ confidence interval for $c^\top\gamma$.

Note that, following Stoye (2009), we can adapt the choice of the quantile to handle near to point-identified cases. Furthermore, we can follow Stoye (2021) to build robust confidence intervals.

H.4 Drawing Points from the Confidence Region for γ

As highlighted in Footnote 27, our tables in Section 8.2 report intervals for the counterfactual outcomes for two reasons. First, a counterfactual outcome must be computed for each parameter value within the confidence region (or estimated identified set) of the parameters. Second, multiple first-stage equilibria can generate several counterfactual

networks for each parameter value and, in turn, multiple counterfactual outcomes.

To address these issues, we proceed as follows. First, we ignore the uncertainty stemming from the second-stage estimates of θ, as it is negligible compared to the uncertainty arising from the first-stage estimates of γ (see Table G.2). Second, rather than using the standard gridding approach—which is computationally intensive—we describe a procedure to draw points uniformly from the confidence region. A similar but simpler procedure applies when drawing points from the estimated identified set. Third, for each draw of the first-stage parameters, we apply an algorithm—detailed in Appendices J.1 and J.2—that produces a probability distribution of equilibrium networks. Using this distribution, we compute the corresponding set of counterfactual outcomes. We report the range of these outcomes (particularly the minimum, median, and maximum) in our tables in Section 8.2. Following Woutersen and Ham (2013), the interval spanned by the minimum and maximum values provides a valid approximation of the confidence interval for the counterfactual outcome under consideration.

In general, drawing points from a confidence region is nontrivial. In a standard GMM framework, one typically evaluates a grid of points in the relevant parameter space and retains those for which the GMM criterion is below the $(1-\alpha)$ quantile. However, in a set-identified context, this approach becomes much more complex. We leverage the specific (smooth) geometry of our identified set to draw points directly from the $(1-\alpha)$ confidence region, as described in detail below.

1. We look for an interior point γ_c in $\widehat{\Gamma}_I^\kappa$. This is known in the convex optimization literature as the Chebyshev center of a polyhedron (Boyd and Vandenberghe, 2004, page 148). Interestingly, it can be solved based on linear programming:

$$\begin{aligned}
&\max_{r\geq 0} r, \\
&\text{s.t. } \frac{1}{\mathrm{M}}\sum_{m=1}^{\mathrm{M}}(-Z_{m,r}B_m^\top\gamma + Z_{m,r}A_m) + r\|\frac{1}{\mathrm{M}}\sum_{m=1}^{\mathrm{M}} Z_{m,r}B_m\|_2 \leq 0, \\
&r = 1,\ldots,\mathrm{R}.
\end{aligned}$$

2. Draw a random direction q on the unit sphere and find the frontier point $\gamma_q = \gamma_c + r_q q$ of $\widehat{\Gamma}_I^\kappa$ ($r_q \geq 0$). Again, this is a linear program.

3. Calculate the outer normal vector of $\widehat{\Gamma}_I^\kappa$ at γ_q. This is the direction q' such that $\delta(q';\widehat{\Gamma}_I^\kappa) = {q'}^\top\gamma_q$. It can be done analytically by calculating the gradient of $g_\kappa(\cdot)$ at γ_q.

4. Calculate the variance $v_\kappa(q')$ of $\mathbb{G}_\kappa(q')$ from (G.2).

5. The point $f_q = \gamma_q + \sqrt{v_\kappa(q')} n_\alpha q'$, i.e., drawn from γ_q in direction q', is a frontier point of the (conservative) confidence region $CR^{\mathrm{M}}_{1-\alpha}$.

6. Draw a norm l uniformly on $[0, 1]$.

7. Pick the point $\gamma_c + l f_q$ which belongs to $CR^{\mathrm{M}}_{1-\alpha}$.

Figure H.4 illustrates the procedure, showing $\gamma_q + 0.8 f_q$ as an example. This process can be repeated S times to generate S draws from the $(1-\alpha)$ confidence region.

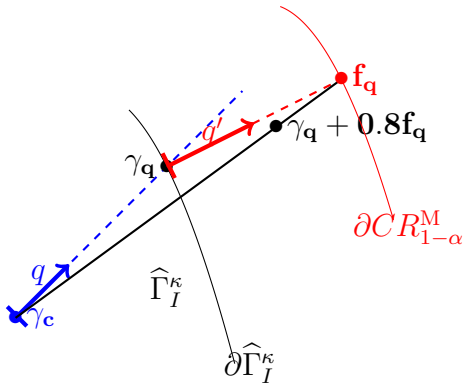


Figure H.4: Drawing points from the confidence region for the first-stage parameters.

I Empirical application

I.1 Results from Demand and Supply

Table I.1 complements Table 6 by presenting the estimated variable profits, prices, marginal costs, and markups at the firm level across different levels of aggregation. For each airline, the first, second, and third rows contain quantities averaged over all products, direct flights and one-stop flights respectively. The fourth and fifth rows contain quantities averaged over direct flights where at least one of the endpoints is a hub, and direct flights where no endpoint is a hub. We can see that airlines charge higher markups on direct flights compared to one-stop flights, which is in line with the fact that consumers prefer to take direct flights (see "*Indirect*" in Table 5, demand panel). The legacy carriers charge higher markups on direct flights where at least one of the endpoints is a hub than on direct flights where no endpoint is a hub, suggesting the existence of a hub premium. This hub premium may be due to the fact that consumers value flying from dense hubs (see "*Nonstop Origin*" in Table 5, demand panel) or to fixed costs due to congestion effects at hubs (see Table 7). While American Airlines, US Airways and Southwest Airlines have lower marginal costs for direct flights, the opposite is true for

Delta and United Airlines.[6] The marginal cost of Southwest Airlines is lower than the marginal cost of the legacy carriers. For direct flights, the difference is quite substantial. For one-stop flights, Southwest Airlines' advantage is small, consistent with the fact that Southwest Airlines uses focus cities rather than hubs. Therefore, the marginal cost savings of offering one-stop flights (see "*Connections*" and "*Indirect*" in Table 5, supply panel) may be less pronounced as not all the features of traditional hubs are used.

Table I.1: Profits by firms.

	Profits (100k)	Price	Marginal cost	Markup	Lerner Index
AA					
All	1.78	453.36	335.20	118.16	0.28
Direct	13.77	402.37	277.42	124.94	0.32
One-stop	0.39	459.26	341.89	117.38	0.27
Direct, hub endpoint	15.06	402.75	276.66	126.09	0.33
Direct, non-hub endpoints	2.00	398.87	284.48	114.40	0.30
DL					
All	1.41	436.45	310.40	126.05	0.31
Direct	12.31	463.26	321.03	142.23	0.33
One-stop	0.33	433.80	309.35	124.45	0.31
Direct, hub endpoint	13.49	482.67	336.83	145.84	0.32
Direct, non-hub endpoints	4.47	334.75	216.44	118.31	0.38
UA					
All	1.25	445.56	328.43	117.13	0.28
Direct	9.17	458.50	334.97	123.53	0.29
One-stop	0.20	443.85	327.56	116.28	0.28
Direct, hub endpoint	11.03	456.82	332.24	124.58	0.29
Direct, non-hub endpoints	2.17	464.88	345.33	119.55	0.29
US					
All	1.30	453.43	336.77	116.67	0.27
Direct	8.99	407.34	275.17	132.17	0.35
One-stop	0.35	459.10	344.34	114.76	0.26
Direct, hub endpoint	10.42	418.96	282.96	136.00	0.35
Direct, non-hub endpoints	3.95	366.22	247.58	118.64	0.36
WN					
All	2.79	419.43	299.51	119.92	0.31
Direct	12.09	365.14	237.09	128.05	0.38
One-stop	0.23	434.40	316.73	117.67	0.29
Direct, hub endpoint	16.49	362.34	233.95	128.39	0.38
Direct, non-hub endpoints	8.88	367.19	239.39	127.80	0.38

Note: Quantities are in USD.

I.2 Estimated Shares of Variable Costs over Operating Costs

To verify if our fixed cost estimates are reasonable, we compute the estimated share of the variable costs over the "operating costs". The former are obtained as marginal costs

[6]Note that the fact that American Airlines, US Airways and Southwest Airlines have lower marginal costs on direct flights does not contradict the negative sign of the coefficient on "*Connections*" in Table 5. In fact, recall that the results in Table 5 should be interpreted *ceteris paribus*. Instead, the results in Table I.1 are obtained by averaging over all itineraries, including those with different characteristics.

times the number of passengers, summed for all direct flights proposed by each airline. The latter are defined as the sum of the variable costs and fixed costs, without considering the congestion costs. Table I.2 reports the minimum and maximum share for each airline. Table I.3 compares these shares with the Federal Aviation Administration (FAA)'s 2018 estimates based on administrative data, revealing similar orders of magnitude.[7]

Table I.2: Estimated shares of the variable costs over the operating costs.

Airline	Min share (in %)	Max share (in %)
AA	72.7 %	83.0 %
DL	69.9 %	81.0 %
UA	68.2 %	79.7 %
US	61.8 %	74.8 %
WN	43.5 %	64.3 %

J Counterfactuals

J.1 Descriptions of the Counterfactual Algorithm

The possibility of multiple PSNE networks raises the question of how to obtain counterfactuals when airlines are allowed to re-optimize their networks and prices. Although the data tell us which equilibrium was played in the past, they do not tell us which equilibrium will be chosen by the players once we change the environment. Previous literature has suggested several ways of solving this problem. For example, the analyst could enumerate all possible equilibria and report some summary measures of the resulting range of counterfactuals (Eizenberg, 2014). Alternatively, the analyst could implement best-response dynamics to select a probability distribution of possible equilibria (Lee and Pakes, 2009; Wollmann, 2018). The first approach is not computationally feasible in our setting, due to the large number of markets and the presence of entry spillovers. Therefore, we follow the second approach. We fix an order of markets and firms.[8] For a given value of the parameters, the first firm in the first market best responds to its competitors in terms of entry and pricing decisions. The second firm similarly best responds, taking into account the best response of the first firm. The third company also best responds, taking into account the best responses of the first and second companies. The algorithm cycles through the firms and markets until no airline wishes to deviate. The procedure

[7] See `https://www.faa.gov/regulations_policies/policy_guidance/benefit_cost`, Section 4 of the Benefit-Cost analysis, Table 4-6.

[8] We focus on the markets where at least one of the endpoints is a hub of the merged entity.

Table I.3: Passenger Air Carriers Filing Schedule P-5.2 Operating and Fixed Costs per Block Hours.

Aircraft Category	Cost per Block Hour								
	Fuel and Oil	Maintenance	Crew	Total Variable	Deprec.	Rentals	Other	Total Fixed	Share Variable
Wide-body more than 300 seats	$5,411	$1,331	$2,356	$9,097	$845	$406	$5	$1,254	87.9%
Wide-body 300 seats and below	$4,080	$1,289	$1,857	$7,227	$685	$366	$8	$1,058	87.2%
Narrow-body more than 160 seats	$2,054	$718	$1,152	$3,925	$355	$217	$10	$582	87.1%
Narrow-body 160 seats and below	$1,741	$737	$1,034	$3,512	$306	$215	$12	$533	86.8%
RJ more than 60 seats	$115	$431	$444	$991	$131	$252	$14	$397	71.4%
RJ 60 seats and below	$92	$479	$470	$1,041	$58	$227	$8	$293	78.0%
Turboprop more than 60 seats	$0	$880	$360	$1,241	$439	$103	$2	$544	69.5%
All Aircraft	$1,681	$727	$1,012	$3,420	$314	$239	$11	$564	85.8%

Source: FAA, `https://www.faa.gov/regulations_policies/policy_guidance/benefit_cost`, Section 4 of the Benefit-Cost analysis, Table 4-6.

is repeated for 50 draws of parameter values from the estimated identified set (or confidence region) of the fixed cost parameters. For each parameter value, we consider four market orderings. In the first ordering, we rank the markets according to which hub is involved, whether the market is served by the merged firm, the size of the merged firm's operations at the endpoints, and the market size (ordering A). In the second ordering, we reverse this ranking (ordering B). In the third and fourth orderings, we rank markets randomly (orderings C and D). For each of the four market orderings, we consider two firm orderings: AA-DL-UA-WL (ordering 1) and the reverse (ordering 2). This procedure generates a distribution of possible equilibria over 400 (i.e., $50 \times 4 \times 2$) counterfactual runs. In the tables of Section 8, we report the minimum, maximum, and median changes in the relevant outcomes under such distribution.

The remainder of the section illustrates the details of the counterfactual algorithm. In particular, we explain the algorithm implemented to simulate the merger under the *Networks vary - without remedies* scenario, given an order of markets and firms and a value of the parameters. The algorithm is structured in the following steps:

1. Latent variables. We determine the realizations of the latent variables that are needed to evaluate the airlines' profits. In particular, from the vector of second-stage estimates, $\hat{\theta}$, we compute the second-stage shocks for each product offered by the airlines before the merger, via BLP inversion. For each airline f, we compute the mean and variance of the second-stage shocks and denote them by μ_f and Σ_f, respectively. When computing μ_f and Σ_f for the merged airline, we consider the second-stage shocks associated with all the products offered by the merging firms before the merger. If both American Airlines and US Airways offer a given itinerary before the merger, then we take the mean value of the second-stage shocks of the two pre-merger products. For each potential product of every airline f, we take 100 random draws from a normal distribution with mean μ_f and variance Σ_f. We store all such draws in a matrix Ξ. Further, for each market $\{a, b\}$ and airline f, we impute the fixed cost shock $\eta_{ab,f}$ as explained in Appendix J.2.

2. Initial state. At the start, all firms except the merged entity are assigned their pre-merger networks and products. The merged entity is assigned the network resulting from combining the pre-merger networks of American Airlines and US Airways. The products initially offered by the merged entity and their observed characteristics are constructed from such a merged network. We denote by $G \coloneqq (G_1, \ldots, G_{\text{N-1}})$ the initial networks of the carriers. We let the firms play the simultaneous pricing game described in Section 3.1, for each draw of the second-stage shocks stored in the matrix Ξ. We save the initial equilibrium prices in a matrix P.

3. Iterations. We take the first firm f in the first market $\{a, b\}$ and let it play its best response as follows. Suppose, for instance, that the initial network G_f is characterized by $G_{ab,f} = 0$. First, we compute airline f's expected variable profits under $(G_{(+ab),f}, G_{-f})$.

To do so, we update the list of products offered by firm f, by adding direct flights between cities a and b. Further, as explained in Appendix D, setting $G_{ab,f} = 1$ induces ripple effects in neighboring markets due to the creation of new products for airline f, changes in the variables "*Nonstop Origin*" and "*Connections*" of existing products of airline f, and changes in the congestion costs of existing products of airline f. Let $\mathcal{M}_{ab,f}$ be the list of markets containing either new products or products with modified covariates of airline f. For each of these products in every market $m \in \mathcal{M}_{ab,f}$, we let airline f find the best-response price, while holding the other prices in P fixed, for every draw of the second-stage shocks stored in the matrix Ξ. We compute airline f's variable profits, average across draws, and get the simulated airline f's expected variable profits, which we denote by

$$\sum_{m \in \mathcal{M}_{ab,f}} \mathbb{E}[\Pi_{m,f}(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G_{(+ab),f}, G_{-f}; \theta) | X^{\oplus}, W^{\oplus}, \text{MS}].$$

We implement a similar procedure to compute the expected variable profits of airline f in each market $m \in \mathcal{M}_{ab,f}$ under G, which we denote by

$$\sum_{m \in \mathcal{M}_{ab,f}} \mathbb{E}[\Pi_{m,f}(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G; \theta) | X^{\oplus}, W^{\oplus}, \text{MS}].$$

Next, we calculate

$$\begin{aligned}\Delta\Pi_{(+ab),f} = &\sum_{m \in \mathcal{M}_{ab,f}} \mathbb{E}[\Pi_{m,f}(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G_{(+ab),f}, G_{-f}; \theta) | X^{\oplus}, W^{\oplus}, \text{MS}] \\ &- \sum_{m \in \mathcal{M}_{ab,f}} \mathbb{E}[\Pi_{m,f}(X^{\oplus}, W^{\oplus}, \text{MS}, \xi^{\oplus}, \omega^{\oplus}, G; \theta) | X^{\oplus}, W^{\oplus}, \text{MS}].\end{aligned}$$

Lastly, we compute

$$\Delta\Pi_{(+ab),f} - \Delta\text{FC}_{(+ab),f}. \tag{J.1}$$

If (J.1) is positive (negative), then the best-response entry of airline f is $G_{ab,f} = 1$ ($G_{ab,f} = 0$). We update G and P and move to the second firm in the first market. We let this firm best respond, while taking into account the first firm's best response. The third firm similarly best responds, while taking into account the first and second firms' best responses, and so on.

4. Stop. We cycle through the firms and markets. When no firm wants to deviate in none of the markets, we stop the procedure. In practice, we have obtained convergence in almost all the cases considered.

Due to computational costs, the above algorithm does not consider all possible entry deviations by each firm. In fact, it imposes that each firm considers adding/deleting direct flights in one market at a time. Nevertheless, at the rest point of the procedure, the

necessary conditions for PSNE that are used in the estimation of the fixed cost parameters hold. Hence, the algorithm provides an equilibrium that is internally consistent with our model. Similar restrictions on the set of admissible deviations are assumed by Eizenberg (2014) and Wollmann (2018).[9]

We also adopt the above algorithm in the merger simulation for the *Networks vary - with remedies* and *Networks vary - PHX dehubbed* scenarios. In the scenario *Networks vary - w/ remedies*, we do not allow the merged entity to exit the markets out of Charlotte, New York, Los Angeles, Miami, Chicago, Philadelphia, and Phoenix that were served before the merger by American Airlines or US Airways. In the scenario *Networks vary - PHX dehubbed* we delete all flights of the merged entity between Phoenix and non-hub cities and do not allow the merged entity to re-enter those markets.

J.2 Imputation of the Fixed Cost Shocks in the Counterfactuals

To perform the counterfactuals, we need to impute the fixed cost shocks. Different approaches have been taken in the literature. For example, Wollmann (2018) draws the fixed cost shocks from a normal distribution with zero mean and variance equal to a fraction of the variance of the systematic fixed costs. Kuehn (2018) finds, for each market, the range of realizations of the fixed cost shocks generating the observed entry/exit patterns and takes the midpoint. We use a procedure that is similar to Kuehn (2018). We repeat the steps below for each value of γ drawn from the estimated identified set at which we run the counterfactual algorithm. When we observe airline f serving market $\{a, b\}$ with direct flights (i.e., $G_{ab,f} = 1$), we infer that this choice must be profitable, giving us an upper bound for $\eta_{ab,f}$. Next, we collect all the markets where airline f does not enter, that are hub markets (non-hub markets) if market $\{a, b\}$ is a hub market (is not a hub market) for airline f, and that face similar congestion costs. These markets give us a vector of lower bounds for $\eta_{ab,f}$. We take the 2.5th percentile of these lower bounds and use it as a lower bound for $\eta_{ab,f}$. Lastly, we set $\eta_{ab,f}$ equal to the midpoint between the lower and upper bounds. We implement a similar procedure to determine the fixed cost shocks for the markets that are not served by airline f in the data. However, instead of the 2.5th percentile, in that case we take the 97.5th percentile to obtain an upper bound. When simulating the merger, the merged entity gets the mean value of the fixed cost shocks imputed to the merging firms by following the above procedure.

[9]The networks at the rest point of our algorithm constitute a pairwise stable outcome, in the sense illustrated in Appendix B. In fact, our algorithm resembles the tâtonnement dynamics discussed by Jackson and Watts (2002), in which agents form or destroy individual connections, taking the remaining network as given and not anticipating future adjustments. Jackson and Watts (2002) show that pairwise stable networks can be achieved by tâtonnement dynamics.

J.3 Additional Tables

Table J.1 reports the percentage change in prices, marginal costs, and markups of American Airlines and the other major airlines. It distinguishes between direct flights and one-stop flights.

Table J.1: Percentage change in prices, marginal cost, and markups.

	Before	**Merger**		
		w/o remedies	w/ remedies	PHX dehubbed
AA: Direct				
Price	406.24	-6.12	-4.78	-5.77
		[-6.49, -5.89]	[-5.21, -4.45]	[-6.45, -5.08]
		[-6.66, -5.79]	[-5.33, -4.44]	[-6.73, -5.08]
Marginal Cost	276.7	-11.31	-9.71	-10.8
		[-12.16, -11.01]	[-10.57, -9.26]	[-12.08, -9.48]
		[-12.41, -10.84]	[-10.67, -9.26]	[-12.5, -9.48]
Markup	129.54	+5.15	+5.79	+4.98
		[+4.78, +5.73]	[+5.68, +6.24]	[+3.99, +5.69]
		[+4.59, +5.95]	[+5.66, +6.24]	[+3.9, +5.73]
Others: Direct				
Price	413.19	+1.48	+0.99	+1.58
		[+1.24, +1.92]	[+0.8, +1.11]	[+1.34, +2.04]
		[+1.03, +2.43]	[+0.69, +1.54]	[+1.32, +2.56]
Marginal Cost	291.6	+1.83	+1.52	+1.7
		[+1.51, +2.33]	[+1.31, +1.73]	[+1.5, +2.35]
		[+1.36, +2.82]	[+1.13, +2.21]	[+1.37, +2.87]
Markup	121.59	+0.73	-0.32	+1.2
		[+0.35, +1.14]	[-0.55, -0.16]	[+0.92, +1.47]
		[+0.23, +1.71]	[-0.77, +0.08]	[+0.52, +1.95]
AA: One-stop				
Price	466.39	-7.36	-4.96	-6.83
		[-8.15, -6.8]	[-5.56, -4.71]	[-7.41, -5.59]
		[-8.15, -6.59]	[-5.58, -4.71]	[-7.66, -5.59]
Marginal Cost	351.28	-12.77	-10.68	-11.92
		[-13.92, -11.8]	[-11.49, -10.3]	[-12.7, -10.1]
		[-13.96, -11.8]	[-11.69, -10.29]	[-12.9, -10.1]
Markup	115.11	+9.22	+12.39	+8.62
		[+8.39, +9.88]	[+12.15, +12.77]	[+7.81, +9.15]
		[+7.23, +10.43]	[+11.55, +13.12]	[+6.55, +9.99]
Others: One-stop				
Price	416.12	+4.67	+4.15	+4.68
		[+4.37, +4.96]	[+4.02, +4.43]	[+4.47, +4.87]
		[+4.37, +5.07]	[+3.95, +4.45]	[+4.45, +5.08]
Marginal Cost	301.18	+6.18	+5.95	+6.15
		[+5.78, +6.53]	[+5.8, +6.32]	[+5.85, +6.39]
		[+5.78, +6.53]	[+5.72, +6.32]	[+5.76, +6.48]
Markup	114.94	+0.77	-0.59	+0.9
		[+0.48, +0.93]	[-0.66, -0.5]	[+0.72, +1.07]
		[+0.41, +1.53]	[-0.79, -0.36]	[+0.39, +1.64]

Note: Percentage changes with respect to thepre-merger scenario are reported.

References

Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *Econometrica*, 60(4):889–917.

Berry, S. T., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Chesher, A. and Rosen, A. M. (2020). Econometric modeling of interdependent discrete choice with applications to market structure. CeMMAP Woking Paper 25/20.

Conley, T. G. (2010). *Spatial Econometrics*, pages 303–313. Palgrave Macmillan UK, London.

Eizenberg, A. (2014). Upstream innovation and product variety in the U.S. home pc market. *The Review of Economic Studies*, 81(3):1003–1045.

Espín-Sánchez, J.-A., Parra, A., and Wang, Y. (2023). Equilibrium uniqueness in entry games with private information. *The RAND Journal of Economics*, 54(3):512–540.

Gualdani, C. (2021). An econometric model of network formation with an application to board interlocks between firms. *Journal of Econometrics*, 224(2):345–370.

Hellmann, T. (2013). On the existence and uniqueness of pairwise stable networks. *International Journal of Game Theory*, 42:211–237.

Jackson, M. O. and Watts, A. (2001). The existence of pairwise stable networks. *Seul Journal of Economics*, 14(3):299–321.

Jackson, M. O. and Watts, A. (2002). The evolution of social and economic networks. *Journal of Economic Theory*, 106(2):265–295.

Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74.

Kojevnikov, D. (2021). The bootstrap for network dependent processes. arXiv Working Papers 101.12312.

Kojevnikov, D., Marmer, V., and Song, K. (2021). Limit theorems for network dependent random variables. *Journal of Econometrics*, 222(2):882–908.

Kuehn, J. (2018). Spillovers from entry: the impact of bank branch network expansion. *The RAND Journal of Economics*, 49(4):964–994.

Lee, R. S. and Pakes, A. (2009). Multiple equilibria and selection by learning in an applied setting. *Economics Letters*, 104(1):13–16.

Leung, M. P. (2023). Network cluster-robust inference. *Econometrica*, 91(2):641–667.

Lewbel, A. (2007). Coherency and completeness of structural models containing a dummy endogenous variable. *International Economic Review*, 48(4):1379–1392.

Mele, A. (2017). A structural model of dense network formation. *Econometrica*, 85(3):825–850.

Molchanov, I. and Molinari, F. (2018). *Random Sets in Econometrics*. Cambridge University Press.

Molinari, F. (2020). Econometrics with partial identification. In *The Handbook of Econometrics*, volume 7a, pages 355–486.

Monderer, D. and Shapley, L. S. (1996). Potential games. *Games and Economic Behavior*, 14(1):124–143.

Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2014). *Lectures on Stochastic Programming: Modelling and Theory*. SIAM.

Sheng, S. (2020). A structural econometric analysis of network formation games through subnetworks. *Econometrica*, 88(5):1829–1858.

Stoye, J. (2009). More on confidence intervals for partially identified parameters. *Econometrica*, 77(4):1299–1315.

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

Topkis, D. M. (1979). Equilibrium points in nonzero-sum n-person submodular games. *SIAM Journal of Control and Optimization*, 17(6):773–787.

Wollmann, T. G. (2018). Trucks without bailouts: Equilibrium product characteristics for commercial vehicles. *American Economic Review*, 108(6):1364–1406.

Woutersen, T. M. and Ham, J. (2013). Calculating confidence intervals for continuous and discontinuous functions of parameters. CeMMAP working papers 23/13, Institute for Fiscal Studies.