

Delta Northwest Fare Analysis

Technical Supplement

Xuange (Adora) Chen · September 2026

A Data provenance and measurement

This supplement accompanies Assessing Fare Changes After the Delta–Northwest Merger. It records the data construction, implementation checks, additional diagnostics, and reconciliation behind the writing sample. The main relative fare contrast is -5.72%. The pre-period evidence limits its causal interpretation; the material below makes the numerical analysis reproducible without treating reproducibility as identification.

The input is the supplied DB1B_2005_2010 copy.csv, containing 4,175,354 aggregate records and 11 fields. Its SHA256 fingerprint identifies the exact file used:

68822d60114c61af9f6ce8d9084aa20276f0d5b1b64a9a80b0e372929f1d19cb

Field	Definition
cr1, cr2	Observed operating-carrier set; segment order is not retained.
yr, qtr	Calendar year and quarter.
cop	0 identifies one coupon; 1 identifies two coupons.
ap1, ap2	Alphabetically ordered airport endpoints; directions combined.
pax	Sampled passenger journeys represented by the aggregate record.
avprc	Mean one-way-equivalent fare in nominal dollars.
nsdst	Airport-pair distance field; zero is not a physical distance estimate.
avdst	Mean itinerary distance; unused in the main estimation.

Table A1. Fields in the supplied extract. Definitions follow Borenstein’s broadened-market documentation. One coupon is an itinerary category and is not relabeled as verified nonstop service.

A row represents a carrier set, unordered airport pair, quarter, and coupon category. The airport directions are already combined. The natural key is cr1, cr2, yr, qtr, cop, ap1, ap2; there are no duplicate natural keys or exact duplicate rows. Missing cr2 is expected in one-coupon records and is not a reason to delete them.

An independent comparison with the NBER Stata archive finds all 4,175,354 records in the same order. Airports and integer-valued fields agree exactly; the maximum difference in mean fare or mean itinerary distance is below 5.01×10^{-12} . However, 42 literal carrier entries NA in the archive appear as blanks in the supplied CSV: 22 in cr1 and 20 in cr2, across 34 records representing 6,453 sampled journeys. No codes are imputed. Those records remain in fare totals, and no passengers with missing carrier labels enter the main sample in 2007. The archive comparison therefore reports the string discrepancies rather than claiming complete file identity.

This audit establishes the relationship between two aggregate files. It does not independently validate the earlier screening of individual DB1B tickets. In particular, the supplied fields cannot recheck cabin, trip purpose, ticket-level outliers, or every upstream itinerary restriction.

B Cleaning and sample construction

Cleaning proceeds in a fixed order. Restrict years to 2005–2010 and quarters to 1–4; this removes no records. Require nonmissing, distinct airports; four same-airport records are removed. Require finite positive pax and avprc; neither condition removes additional records. The clean file thus contains 4,175,350 records. No new fare trimming, winsorization, carrier-family imputation, passenger expansion, or filling of unobserved quarters is performed. The full attrition sequence is recorded in row_cleaning_flow.csv.

For each retained record i , form $\text{pax}_i \times \text{avprc}_i$. For airport pair r and quarter t , define fare as the sum of those products divided by the sum of pax, then take its natural logarithm. This reconstructs the passenger-weighted mean represented by the aggregate records. The code calls the numerator revenue, but it is a sampled-fare construction, not measured airline revenue. Assertions check conservation of passengers and this numerator, unique route-quarter keys, and finite log outcomes.

The clean panel contains 39,505 airport pairs, 609,396 observed route-quarters, and 229,384,433 sampled journeys. Every airport pair already has alphabetically ordered endpoints. The quarter index is $t = (\text{yr} - 2005) \times 4 + \text{qtr}$, so 2008Q4 is $t = 16$. Fares are nominal one-way-equivalent dollars. A common quarterly deflator is absorbed by quarter effects in a log-fare model; it does not control for heterogeneous route responses to economic shocks.

Main eligibility uses only 2007: a route must have observations in all four quarters and at least 100 sampled journeys in each, leaving 5,535 eligible routes. For a given airline code, single-carrier traffic comprises $\text{cop} = 0$ records with that cr1 and $\text{cop} = 1$ records with that code in both fields. Its denominator is all passengers on the airport pair, including mixed-carrier journeys.

An overlap route has both DL and NW single-carrier shares of at least 5% in the same quarter in at least three 2007 quarters. A comparison route has no DL or NW code in either field anywhere in 2007 and a combined single-carrier share of at least 5% for AA, AS, CO, UA, US, and HP in at least three quarters. These rules select 156 overlap and 105 comparison routes; 5,274 other eligible routes are outside the main comparison. The thresholds are documented analysis choices, without a claim of historical preregistration.

All available 2005–2010 observations for the selected routes enter estimation: 6,259 route-quarters, constructed from 108,639 input records and 9,220,845 sampled journeys. Of the theoretical 6,264 cells, five comparison-route cells are missing. CLD–LAX appears in 20 quarters and LAS–PSP in 23; all 156 overlap routes have 24 quarters. The balanced check keeps 259 routes and 6,216 cells, but its completeness rule uses post-merger availability.

Zero distance is retained in fare construction because distance is not the outcome or a baseline control. SJU–STT is the only main-sample route with zero recorded distance, in all 24 quarters. Omitting it yields -5.58% . A zero is not treated as a physical distance. Within-cell ticket-fare variances are unavailable, so dispersion across avprc records is not used as evidence about individual-fare dispersion or coordination.

C Estimation and numerical validation

The baseline model uses y , the log of the passenger-weighted route-quarter fare, route effects α , quarter effects λ , and the overlap-by-post interaction. Post begins in 2008Q4; the route classification remains fixed throughout the panel.

$$y_{rt} = \alpha_r + \lambda_t + \beta(T_r \times Post_t) + \varepsilon_{rt}$$

The baseline gives each observed route-quarter equal weight. The fixed-passenger-weight check instead gives a route its average 2007 pax in every observed quarter. The latter changes the weighted contrast while keeping the fare construction fixed. Every reported percentage is $100[\exp(\beta) - 1]$; interval endpoints receive the same transformation.

The implementation removes route means from the outcome and every regressor, including all quarter indicators. Let M denote the projection that removes route effects, D the quarter-indicator matrix, and z the overlap-by-post interaction. Estimation fits $My = (Mz)\beta + (MD)\lambda + u$. Removing route means from y and z alone is generally not equivalent on an unbalanced panel. The weighted specification uses weighted route means and the corresponding weighted least-squares transformation.

For the transformed design X , the cluster covariance is the ordinary sandwich using route-level score sums: $(X'X)^{-1}[\sum g \text{ sg sg}'](X'X)^{-1}$, where $\text{sg} = Xg'ug$. It is multiplied by the CR1 correction $G/(G - 1) \times (N - 1)/(N - G - K)$. G is the number of route clusters, N the number of observed cells, and K the number of regressors after absorbing route effects. The main model has $G = 261$, $K = 24$, and 5,974 residual degrees of freedom. Confidence intervals and two-sided p -values use $t(G - 1)$.

The baseline coefficient is -0.0589, with clustered standard error 0.0189 and $p = 0.0020$. Its transformed contrast is -5.72%, with 95% interval [-9.16%, -2.15%]. Route clustering allows dependence within an airport pair but does not allow unrestricted correlation across pairs sharing airports or airline networks. The interval does not incorporate bias from an unsuitable comparison group.

Independent regressions with explicit route and quarter indicators verify the absorbed implementation. The actual-data subset contains 26 routes and 619 cells, including both incomplete main-sample routes. For unweighted and weighted pooled models, coefficient and clustered-standard-error differences are below 10^{-9} . A separate event regression checks all dynamic coefficients and standard errors, with maximum differences below 2.4×10^{-14} . Legacy-model checks separately confirm the conventions needed to reproduce the supplied original code. These are implementation checks, not evidence for causal assumptions.

The main analysis uses no random draw. Outputs retain input fingerprints, software versions, route classifications, estimates, covariance-based tests, and numerical checks. Figure points come from saved estimates. The model is rerunnable from the original aggregate CSV without manually editing intermediate data or figure values.

D Complete event study estimates

The event model replaces the single interaction with one overlap-by-quarter interaction for every quarter except 2008Q3. It uses all 24 quarters, with route and quarter effects. Table A2 supplies the numerical values behind the writing sample's Figure 1. The zero reference has no estimated standard error.

Quarter	Log-fare contrast	SE	95% interval
2005Q1	+0.0097	0.0307	[-0.0508, +0.0701]
2005Q2	-0.0265	0.0326	[-0.0907, +0.0377]
2005Q3	-0.0208	0.0318	[-0.0834, +0.0418]
2005Q4	+0.0185	0.0293	[-0.0393, +0.0762]
2006Q1	-0.0343	0.0293	[-0.0919, +0.0233]
2006Q2	+0.0450	0.0250	[-0.0042, +0.0941]
2006Q3	+0.0822	0.0246	[+0.0338, +0.1307]
2006Q4	+0.0545	0.0244	[+0.0064, +0.1026]
2007Q1	+0.0154	0.0245	[-0.0328, +0.0636]
2007Q2	+0.0492	0.0225	[+0.0050, +0.0935]
2007Q3	+0.0385	0.0193	[+0.0005, +0.0766]
2007Q4	+0.0478	0.0174	[+0.0135, +0.0820]
2008Q1	+0.0367	0.0176	[+0.0022, +0.0713]
2008Q2	+0.0330	0.0102	[+0.0129, +0.0530]
2008Q3	0 (reference)	—	—
2008Q4	-0.0383	0.0113	[-0.0605, -0.0162]
2009Q1	-0.0825	0.0209	[-0.1237, -0.0412]
2009Q2	-0.0334	0.0218	[-0.0764, +0.0096]
2009Q3	-0.0118	0.0207	[-0.0526, +0.0290]
2009Q4	+0.0009	0.0196	[-0.0377, +0.0396]
2010Q1	-0.0425	0.0208	[-0.0835, -0.0014]
2010Q2	-0.0153	0.0208	[-0.0562, +0.0257]
2010Q3	-0.0645	0.0256	[-0.1149, -0.0142]
2010Q4	-0.0331	0.0225	[-0.0774, +0.0113]

Table A2. Log-fare contrasts relative to 2008Q3. Intervals are pointwise 95% intervals using route-clustered standard errors and $t(260)$. They are not simultaneous confidence bands or estimates of a causal response to completion. $N = 6,259$; 261 airport pairs.

The joint pre-completion test is $F(14, 260) = 8.925$, $p = 7.09\text{e-}16$. A separate model restricted to 2005–2007, omitting 2007Q4, gives $F(11, 260) = 7.456$, $p = 4.11\text{e-}11$. Both concern changes in observed relative gaps, potentially including differential seasonality. The second test avoids an announcement-period reference.

E Additional diagnostics

Figure A1 displays the underlying mean log-fare paths. Each quarterly mean gives observed routes equal weight within its group. The differences in levels are consistent with the route characteristics reported in the writing sample. Fixed effects absorb persistent level differences, but the plot also allows the reader to examine changes in the gap across quarters.

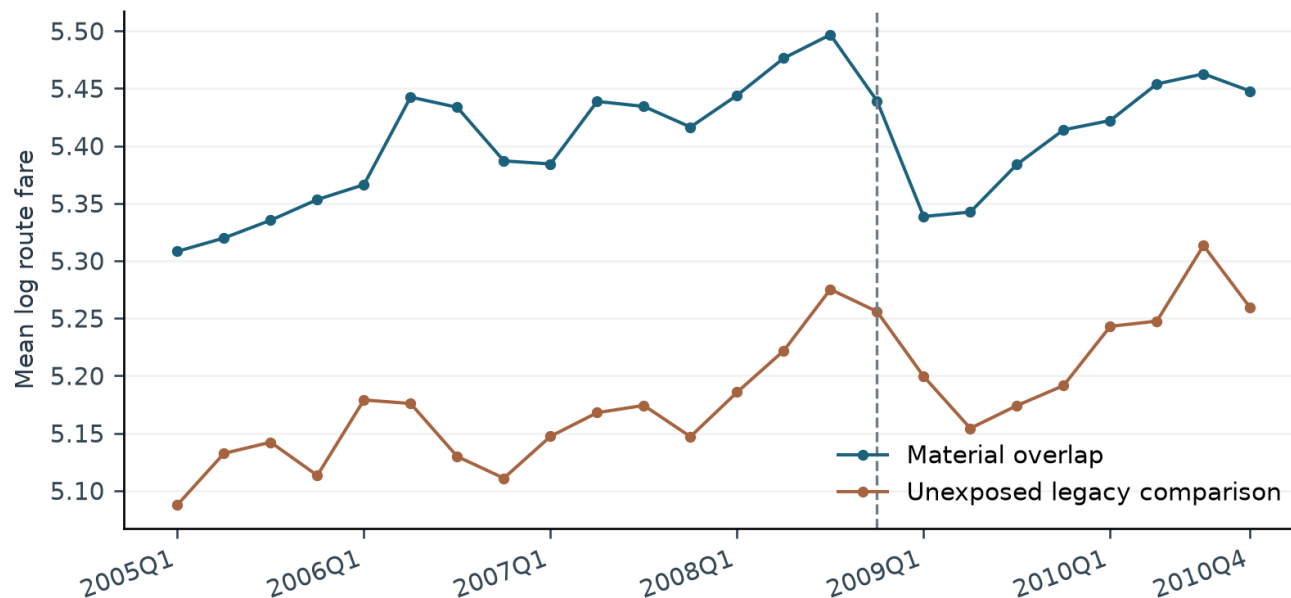


Figure A1. Descriptive mean log fares. The dashed line marks completion in 2008Q4. The plotted means use the main route classification and equal route weights. Missing route-quarters remain missing.

A supplementary nearest-neighbor exercise matches each of the 156 overlap routes to one of the 105 comparison routes, allowing replacement. Features are log average 2007 fare, log average 2007 pax, the supplied distance field divided by 1,000, and the mean 2007 one-coupon passenger share. Each feature is centered and divided by its sample standard deviation across the 261 candidate routes; Euclidean distance selects the nearest comparison. Treated routes receive weight one, and a selected comparison receives its reuse count.

This procedure selects only 28 distinct comparison routes, with maximum reuse of 30. The matched panel has 4,416 cells and 184 routes. Its fare contrast is -4.55%, but the pre-completion event coefficients remain jointly significant: $F(14, 183) = 12.278$, $p < 0.001$. Reuse concentration and remaining pre-period differences make this an informative diagnostic, not a solution to the comparison problem. The inherited distance feature also retains the source zero sentinel, which further limits its interpretation as a validated match on distance.

The writing sample's other checks restrict to complete panels, omit announcement or recession quarters, remove the zero-distance route, or change fixed regression weights. Each addresses the stated numerical sensitivity. None estimates the unobserved counterfactual by itself. In particular, omitting 2008Q3–2009Q4 changes the before-and-after periods and does not isolate a recession-free merger effect.

F Alternative exposure definitions

The four contrasts below use the same passenger-weighted fare construction and fixed-effects implementation. They change route classification and eligibility, and consequently concern different populations.

Route definition	Overlap routes	Comparison routes	Change %
Material overlap in 2007	156	105	-5.72
Any code in 2007	3,706	1,379	-1.23
Before completion / legacy controls	8,180	6,424	+0.80
Before completion / all controls	8,180	23,664	+1.54

Table A3. Alternative route populations. Full coefficient estimates, clustered standard errors, intervals, and observation counts are in `regression_results.csv`. The main rules are stated in Section B.

The 2007 any-code definition retains routes with at least 100 sampled journeys in each 2007 quarter. A route is treated if DL appears in either carrier field at least once in 2007 and NW appears at least once that year. The appearances need not be in the same quarter or in separate single-carrier records. Controls are other eligible routes with positive single-carrier traffic from the stated legacy group during 2007; they may include one merging carrier. This yields 3,706 treated routes, 1,379 controls, and 121,998 cells, with a -1.23% contrast.

The broad pre-completion definition first keeps routes observed in at least four quarters anywhere in 2005–2010. A route must also be observed before 2008Q4 to be classified. Treatment requires any DL and any NW appearance before that quarter. The legacy-control variant keeps other such routes with positive single-carrier legacy traffic before completion: 8,180 treated routes, 6,424 controls, and 332,949 cells, yielding +0.80%. The all-control variant instead uses every other classified route: 8,180 treated routes, 23,664 controls, and 594,407 cells, yielding +1.54%.

These broad rules use service information after the public announcement and a four-quarter availability rule that can use post-completion observations. They do not require material, sustained, contemporaneous, or independently operated overlap. Their results should therefore not be averaged with the main estimate or selected according to a preferred sign.

Two additional variants help isolate aspects of these changes. Pairing 2007 any-code treatment with the original strict 105-route comparison group gives -5.43% across 91,445 cells. Requiring both pre- and post-completion observations in the broad legacy sample gives +0.80% across 331,606 cells. These variants are retained in the output registry so that the reader can distinguish exposure changes from comparison and availability changes.

Operating codes, ticketing carriers, marketing carriers, and parent-airline affiliations answer different economic questions. A dated affiliation crosswalk and finer itinerary records would support a more defensible competition measure. Ticket-level prices, capacity, schedules, quality, and demand information would be needed to connect a fare contrast to specific competitive mechanisms or consumer welfare.

G Reconciliation with the original implementation

Table A4 changes one analytical component at a time. It separates implementation, fare measurement, eligibility, treatment classification, and comparison selection. The sequence is an accounting exercise: a different order can change the size attributed to each step.

Specification	Change %	Overlap	Comparison	N
A Original implementation	+5.13	12	32,214	596,497
B Correct fixed effects only	+5.02	12	32,214	596,497
C Passenger-weighted fare	+1.97	12	32,214	596,497
D Add 2007 eligibility	+0.48	12	5,523	132,430
E Change overlap classification	-0.85	156	5,379	132,430
F Restrict comparison routes	-5.72	156	105	6,259

Table A4. Sequential reconciliation. Change is $100[\exp(\beta) - 1]$; N counts route-quarters. Row A reproduces the supplied original code. Rows B–F use correct route and quarter effects. Inference conventions also change from the legacy normal-reference procedure to the stated t-reference CR1 procedure; the coefficient change from A to B comes from the estimating equation.

The original implementation takes an unweighted average of aggregate-record fares, uses a hand-coded list that selects 12 treated airport pairs, and compares those pairs with all other retained routes. It does not implement a documented comparable-legacy-route screen. The original coefficient is reproducible: +5.13% using the exact exponential conversion. Correcting only the fixed-effects transformation yields +5.02%. That correction alone does not reverse the sign.

Row C changes only the fare measure on the same observations and groups. Row D adds the 2007 eligibility rule while retaining the old labels. Row E uses empirical material overlap within the same eligible pool. Row F retains the 156 empirical overlap routes and restricts comparisons to the 105 unexposed legacy routes. The final –5.72% contrast therefore reflects changes in measurement and population as well as implementation.

The supplied original event code reproduces the point pattern and error bars in the original Figure 2. However, its coefficients and standard errors do not match any of the 16 entries printed in the original Table B1 at the table’s precision. Its event window contains 418,297 observations, rather than the table’s 596,000. The supplied script does not export that table, so its historical source remains unverified. The replication files retain the transcription and the coefficient-by-coefficient comparison without attributing a motive to the discrepancy.

The old event specification also differs from its stated interpretation. It demeans the outcome over the full 24-quarter sample before restricting to a 17-quarter window, but does not apply the corresponding route transformation to the event indicators or reintroduce route effects. Omitting 2008Q3 therefore does not make every other coefficient the stated change relative to that quarter. The new event study uses the complete specification in Sections C–D, and the writing sample does not splice coefficients from the two versions.

The original dispersion regression is retained only for audit. A variance across aggregate-record means omits within-record variation and cannot recover ticket-level price dispersion. Neither that calculation nor the sign of the new mean-fare contrast establishes a coordination mechanism.

H Literature and replication materials

Luo (2014) studies the Delta–Northwest transaction and reports limited fare increases on overlap airport pairs. The present exercise does not claim to overturn that finding: its aggregate outcome, service criteria, comparison routes, and estimating sample must be assessed on their own terms. Orchinik and Remer (2023) show why potentially affected control routes and the choice of comparison method matter for airline-merger estimates. Their later related paper, Remer and Orchinik (2026), further motivates attention to pre-merger fare trends and effects extending beyond overlap routes. A non-overlap route should not be presumed untreated solely from its label.

Roth (2022) explains why pre-period tests cannot certify parallel trends and why conditioning inference on passing such tests creates difficulties. Here, the observed rejections are themselves informative diagnostics; no specification is promoted to a causal estimate because it happens to pass a selected test. The principal contribution is a transparent comparison and an audit of what the evidence can support.

The public repository contains the writing sample, this supplement, the supplied aggregate dataset, analysis code, saved outputs, and source-audit records: [Source](#). The repository README gives the current download and run instructions. The main entry point is `analysis/run_analysis.py`. It reads the input CSV and creates the panels, regressions, figures, diagnostics, and original-code reconciliation; it does not overwrite the input. The optional `provenance/verify_source_identity.py` compares the input with the separately downloaded NBER source archive. Document source files and the builder link reported numbers to saved results.

References

Borenstein, Severin. n.d. Description of “Market Data” Files Created by Severin Borenstein. [Source](#)

Delta Air Lines. 2008a. Delta Air Lines, Northwest Airlines Combining To Create America’s Premier Global Airline. April 14. [Source](#)

Delta Air Lines. 2008b. Delta and Northwest Merge, Creating Premier Global Airline. October 29. [Source](#)

Luo, Dan. 2014. The Price Effects of the Delta/Northwest Airline Merger. *Review of Industrial Organization* 44 (1): 27–48. [Source](#)

National Bureau of Economic Research. n.d. Department of Transportation DB1A/DB1B. [Source](#)

Orchinik, Reed, and Marc Remer. 2023. What’s the Difference? Measuring the Effect of Mergers in the Airline Industry. Working paper, September 6 version. [Source](#)

Remer, Marc, and Reed Orchinik. 2026. Multimarket Contact and Prices: Evidence From an Airline Merger Wave. MIT Sloan Research Paper 7158-24, June 4 version. [Source](#)

Roth, Jonathan. 2022. Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends. *American Economic Review: Insights* 4 (3): 305–322. [Source](#)

Sources and literature checked September 21, 2026. The 2023 and 2026 working papers are cited as separate dated versions. Source-archive fingerprints and field-level discrepancies are retained in the provenance records.