

L'impact de la congestion aéroportuaire sur la durée des vols en Europe

Une analyse économétrique sur données de panel

Fadli Aaron - Université de Tours - Master 1 MECEN

04 janvier 2026

Abstract

Cette étude examine l'impact de la congestion aéroportuaire sur la durée effective des vols en Europe, en exploitant les données OPDI (Operational Performance Data Inputs) d'Eurocontrol pour la période mai-juillet 2023. À l'aide d'une approche économétrique en données de panel au niveau route-jour, nous estimons l'effet causal de la congestion au départ sur la durée moyenne des vols. Nos résultats, basés sur un modèle à effets fixes bidirectionnels (route et temps), révèlent qu'une augmentation de 10 départs quotidiens à l'aéroport de départ est associée à un allongement significatif de la durée de vol. Cette analyse contribue à la littérature sur l'efficacité opérationnelle du transport aérien et fournit des éléments empiriques utiles pour la régulation et la gestion de la capacité aéroportuaire.

Contents

1	Présentation de la question de recherche	3
1.1	Contexte et motivation	3
1.2	Question de recherche et hypothèses	3
2	Présentation des données	4
2.1	Source et périmètre	4
2.1.1	Base OPDI d'Eurocontrol	4
2.1.2	Période d'étude	4
2.2	Méthodologie de collecte et traitement	4
2.3	Nettoyage et construction des variables	4
2.3.1	Filtres appliqués	4
2.3.2	Agrégation en panel route-jour	4
2.4	Statistiques descriptives de l'échantillon	5
3	Statistiques descriptives	6
3.1	Distribution des variables clés	6
3.1.1	Durée des vols	6
3.1.2	Congestion aéroportuaire	7
3.2	Analyse de la corrélation	7
3.3	Évolution temporelle	8
3.4	Décomposition de variance (Between vs. Within)	8
4	Présentation des spécifications économétriques et résultats	9
4.1	Cadre méthodologique	9
4.1.1	Modèle de base	9
4.1.2	Équation générale	9
4.1.3	Stratégies d'identification	9
4.2	Résultats des estimations	10

4.3	Interprétation économique	11
4.3.1	Effet marginal	11
4.3.2	Limites de l'interprétation causale	11
4.4	Tests de spécification	12
5	Choix du modèle préféré et justification	12
5.1	Arbitrage entre les spécifications	12
5.1.1	Critères de sélection	12
5.1.2	Analyse comparative	12
5.2	Modèle préféré : FE two-way (route + date)	13
5.3	Tests de robustesse	13
5.3.1	Spécification logarithmique	13
5.3.2	Spécification par quintiles	13
5.4	Conclusion sur le choix de spécification	14
6	Conclusion générale	15
7	Annexes	15
7.1	Annexe A : Code R complet	15
7.2	Annexe B : Statistiques descriptives complémentaires	15

1 Présentation de la question de recherche

1.1 Contexte et motivation

Le transport aérien européen connaît une croissance soutenue de son trafic depuis plusieurs décennies, avec une reprise particulièrement forte suite à la pandémie de COVID-19. Cette dynamique pose des défis importants en termes de gestion de capacité aéroportuaire et d'encombrement. La congestion aéroportuaire, définie comme la saturation temporaire des infrastructures d'un aéroport (pistes, taxiways, gates), génère des externalités négatives significatives : retards, surconsommation de carburant, coûts opérationnels accrus et émissions de gaz à effet de serre supplémentaires.

D'un point de vue économique, la congestion reflète un déséquilibre entre l'offre de capacité aéroportuaire et la demande de créneaux horaires. Lorsque le nombre de mouvements d'avions s'approche de la capacité maximale, les temps d'attente au sol et en vol augmentent, ce qui se traduit par une dégradation de la performance opérationnelle. Cette problématique revêt une importance particulière en Europe, où plusieurs hubs majeurs (Londres-Heathrow, Paris-Charles de Gaulle, Amsterdam-Schiphol) opèrent régulièrement proche de leur capacité maximale.

1.2 Question de recherche et hypothèses

Notre question de recherche centrale est la suivante :

Quel est l'impact causal de la congestion aéroportuaire au départ sur la durée effective des vols en Europe ?

Cette question s'inscrit dans la littérature empirique sur l'efficacité opérationnelle du transport aérien et les effets de la congestion. Nous formulons les hypothèses suivantes :

Hypothèse 1 (H1) : *La congestion au départ allonge significativement la durée des vols*, en raison de délais au roulage, de temps d'attente en file d'attente au décollage, et potentiellement de routages moins directs imposés par le contrôle aérien.

Hypothèse 2 (H2) : *L'effet de la congestion persiste après contrôle des effets fixes de route*, ce qui suggère que la variation temporelle de la congestion a un impact causal distinct des caractéristiques structurelles des routes.

Hypothèse 3 (H3) : *L'effet est non-linéaire*, avec un impact marginal croissant aux niveaux élevés de congestion.

2 Présentation des données

2.1 Source et périmètre

2.1.1 Base OPDI d'Eurocontrol

Les données proviennent de la base **OPDI (Operational Performance Data Inputs)** mise à disposition par Eurocontrol, l'organisation européenne pour la sécurité de la navigation aérienne. Cette base compile les informations de vol transmises par les compagnies aériennes et les services de navigation aérienne pour l'ensemble de l'espace aérien européen.

Caractéristiques de la base :

- **Granularité** : Niveau vol individuel
- **Fréquence** : Données quotidiennes
- **Couverture géographique** : Espace aérien européen (codes OACI E, L, U)
- **Variables clés** : Aéroports de départ et d'arrivée (ADEP/ADES), heures de première et dernière détection radar (first_seen/last_seen), date de vol (DOF)
- **Format** : Fichiers Parquet mensuels

2.1.2 Période d'étude

Nous retenons la période **mai-juillet 2023**, soit trois mois consécutifs de la saison d'été. Ce choix se justifie par :

1. **Représentativité saisonnière** : La période estivale connaît les volumes de trafic les plus élevés, maximisant la probabilité d'observer des épisodes de congestion.
2. **Stabilité opérationnelle** : Absence de grèves majeures ou d'événements exceptionnels ayant affecté le trafic durant cette période.
3. **Contrainte de volume** : Trois mois représentent un compromis entre richesse des données (plus d'1,3 million de vols) et faisabilité technique du traitement.

2.2 Méthodologie de collecte et traitement

Justification de l'approche : L'utilisation de fonctions dédiées avec gestion d'erreurs robuste garantit la reproductibilité de la collecte et évite les téléchargements redondants (vérification d'existence des fichiers). Le format Parquet, optimisé pour le stockage en colonnes, permet une lecture efficace des données massives.

2.3 Nettoyage et construction des variables

2.3.1 Filtres appliqués

Justification des filtres :

1. **Restriction géographique** : Garantit l'homogénéité de l'espace aérien (régulation commune) et évite les biais liés à des régimes opérationnels très différents (vols intercontinentaux).
2. **Validation temporelle** : Élimine les observations aberrantes ou incomplètes (erreurs de transmission, vols annulés).
3. **Durée de vol** : Calculée comme la différence entre la dernière et la première détection radar, ce qui inclut le temps en vol effectif ainsi que les phases de roulage et d'attente au sol. C'est cette durée qui capture l'impact de la congestion.

2.3.2 Agrégation en panel route-jour

Choix méthodologiques :

1. **Niveau d'agrégation route-jour** : Permet de contrôler l'hétérogénéité des routes (distance, profil opérationnel) via des effets fixes route, tout en exploitant la variation temporelle de la congestion.
2. **Mesure de congestion** : Le nombre total de départs quotidiens à l'aéroport de départ (`congestion_dep_day`) est une proxy directe de la pression sur les infrastructures au sol et en phase initiale de vol.
3. **Seuil de 20 observations** : Les routes observées moins de 20 jours sont exclues pour garantir une variation temporelle suffisante dans l'estimation des effets fixes.

2.4 Statistiques descriptives de l'échantillon

Table 1: Dimensions du panel route-jour

Nombre de routes	Nombre de jours	Observations totales
9 309	92	507 552

Table 2: Statistiques descriptives des variables principales

Variable	Moyenne	Écart-type	Minimum	Médiane	Maximum
Durée moyenne (min)	99.89	53.10	1.08	90.08	724.75
Congestion départ (vols/jour)	140.29	118.68	1.00	107.00	518.00
Vols route-jour	2.41	2.40	1.00	1.00	31.00

Interprétation :

- Le panel comprend 9,309 routes observées sur 92 jours, pour un total de 507,552 observations, ce qui permet d'exploiter la variation temporelle intra-route dans un cadre à effets fixes.
- Les statistiques descriptives montrent une forte hétérogénéité des routes, tant en termes de durée des vols que de congestion au départ.
- Le nombre moyen de vols par route-jour, légèrement supérieur à deux, indique une agrégation de plusieurs vols individuels, limitant le bruit de mesure.

3 Statistiques descriptives

3.1 Distribution des variables clés

3.1.1 Durée des vols

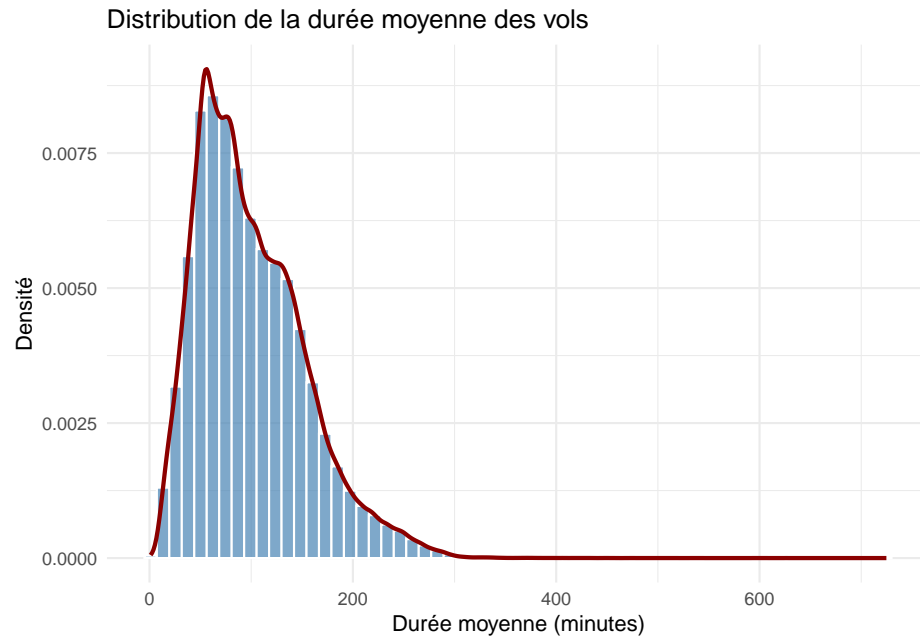


Figure 1: Distribution de la durée moyenne des vols (en minutes)

La distribution de la durée moyenne des vols montre une légère asymétrie positive, avec une concentration autour de 90-120 minutes (vols court et moyen-courrier intra-européens). La queue de distribution vers les durées élevées correspond aux liaisons vers les périphéries de l'espace européen (par exemple, vers les îles Canaries ou Chypre).

3.1.2 Congestion aéroportuaire

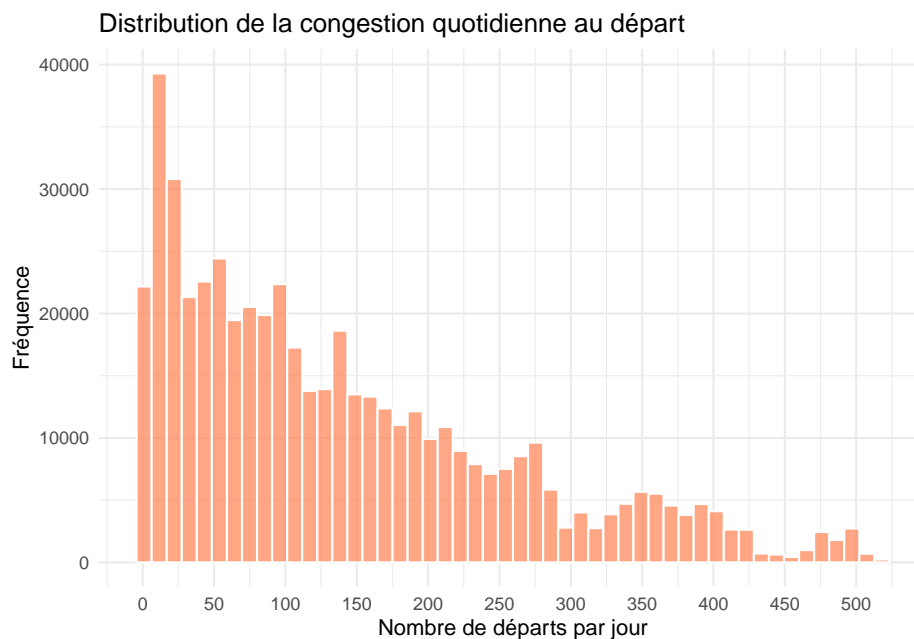


Figure 2: Distribution de la congestion quotidienne au départ (nombre de vols)

La distribution de la congestion est fortement asymétrique, avec une majorité d’aéroports présentant des niveaux modérés et quelques hubs très actifs présentant des volumes quotidiens élevés. Cette hétérogénéité justifie l’inclusion d’effets fixes route pour contrôler les différences structurelles entre aéroports.

3.2 Analyse de la corrélation

Table 3: Matrice de corrélation des variables principales

	duration_mean	congestion_dep_day	nb_flights_route_day
duration_mean	1.000	0.097	-0.114
congestion_dep_day	0.097	1.000	0.338
nb_flights_route_day	-0.114	0.338	1.000

Observations :

- La corrélation entre la durée moyenne des vols et la congestion au départ est positive mais de faible ampleur (0.097), suggérant un effet brut limité de la congestion.
- Une corrélation plus marquée est observée entre la congestion et le nombre de vols par route-jour (0.338), indiquant que les routes les plus fréquentées sont associées à des aéroports plus congestionnés.
- Enfin, la corrélation négative entre la durée des vols et le nombre de vols (-0.114) reflète la coexistence de routes courtes à forte fréquence et de routes plus longues moins denses.
- Ces corrélations brutes ne capturent pas les effets de composition (routes longues vs. courtes, hubs vs. régionaux), d’où la nécessité d’une approche économétrique en panel.

3.3 Évolution temporelle

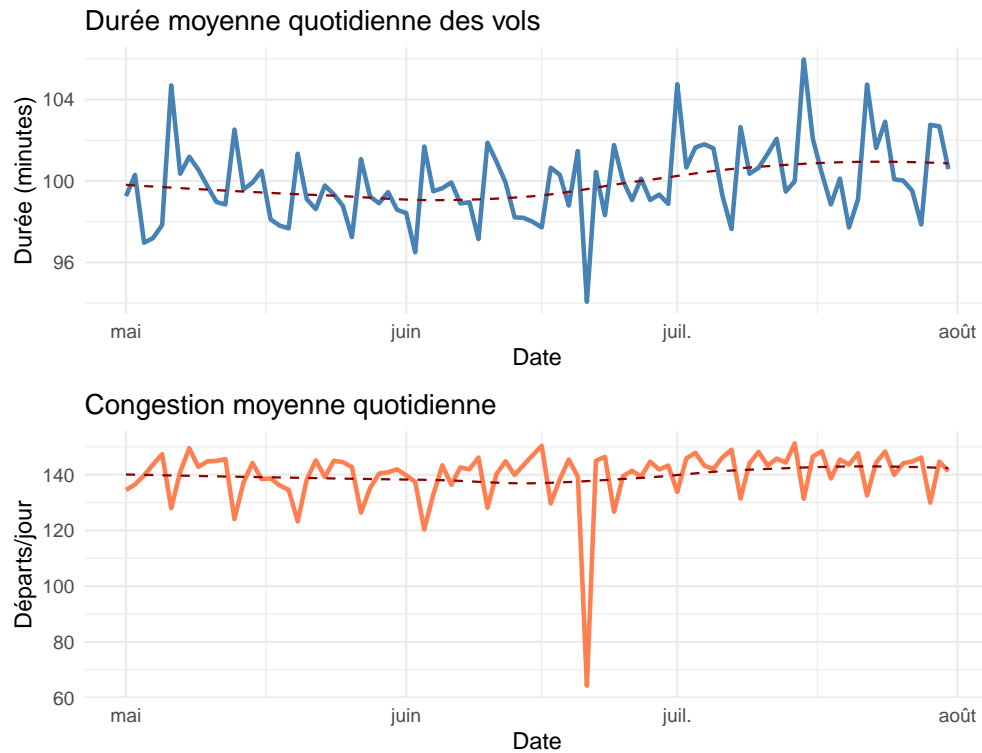


Figure 3: Évolution temporelle de la durée moyenne et de la congestion

Interprétation : Les séries temporelles mettent en évidence une variabilité intra-période et une légère évolution des niveaux moyens de congestion et de durée des vols au cours de la période étudiée, notamment en juillet. Cette variation temporelle agrégée, susceptible de refléter des chocs communs affectant l'ensemble des routes, justifie l'inclusion d'effets fixes temporels dans les spécifications à effets fixes bidimensionnels.

3.4 Décomposition de variance (Between vs. Within)

Table 4: Décomposition de la variance : Between vs. Within

Variable	Var. totale	Var. inter-routes (Between)	Var. intra-route (Within)	Part Within (%)
Durée moyenne	2 820.0	2 680.9	256.7	9.1
Congestion départ	14 084.7	12 489.3	413.0	2.9

Implications :

La décomposition de la variance montre que l'hétérogénéité inter-routes explique l'essentiel de la variabilité observée. Pour la durée moyenne des vols, la variance between (2,681) représente la majeure partie de la variance totale (2,820), la composante within ne comptant que pour 9.1 %. De même, pour la congestion au départ, la part intra-route est très limitée (2.9 %), indiquant que les différences persistantes entre routes dominent la variation observée.

4 Présentation des spécifications économétriques et résultats

4.1 Cadre méthodologique

4.1.1 Modèle de base

Notre variable dépendante est la **durée moyenne du vol** pour une route i à la date t , notée $Duration_{it}$. La variable explicative d'intérêt est le **niveau de congestion au départ** $Congestion_{it}$, mesuré par le nombre total de départs quotidiens à l'aéroport de départ. Nous estimons différentes spécifications pour tester la robustesse de l'effet estimé.

4.1.2 Équation générale

$$Duration_{it} = \beta_0 + \beta_1 Congestion_{it} + \beta_2 NbFlights_{it} + \mathbf{X}'_{it}\gamma + \alpha_i + \delta_t + \varepsilon_{it}$$

où :

- $Duration_{it}$: durée moyenne (en minutes) pour la route i au jour t
- $Congestion_{it}$: nombre de départs à l'aéroport de départ au jour t
- $NbFlights_{it}$: nombre de vols sur la route i au jour t (contrôle pour la fréquence de la route)
- \mathbf{X}_{it} : vecteur de contrôles calendaires (jour de la semaine, mois)
- α_i : effet fixe route (absorbe l'hétérogénéité invariante dans le temps : distance, profil opérationnel)
- δ_t : effet fixe temporel (absorbe les chocs communs à toutes les routes : météo, grèves, saisonnalité)
- ε_{it} : terme d'erreur

Le coefficient d'intérêt β_1 mesure l'**effet marginal** d'un départ supplémentaire sur la durée moyenne, toutes choses égales par ailleurs.

4.1.3 Stratégies d'identification

Nous comparons quatre spécifications pour évaluer la sensibilité des résultats :

1. **Modèle poolé (OLS)** : Estimation par les moindres carrés ordinaires sans effets fixes. Sert de benchmark mais souffre de biais d'hétérogénéité non observée.
2. **Modèle à effets fixes route (Within)** : Contrôle l'hétérogénéité invariante dans le temps (distance, profil de route). Identification repose sur la variation temporelle de la congestion au sein de chaque route.
3. **Modèle à effets aléatoires (RE)** : Hypothèse que α_i est non corrélé avec les régresseurs. Moins robuste mais potentiellement plus efficace si l'hypothèse est vérifiée.
4. **Modèle à effets fixes bidirectionnels (Two-way FE)** : Inclut à la fois des effets fixes route (α_i) et des effets fixes temps (δ_t). Spécification la plus exigeante, contrôle simultanément l'hétérogénéité de route et les chocs temporels communs.

4.2 Résultats des estimations

Table 5: Effet de la congestion au départ sur la durée moyenne des vols

	<i>Dependent variable:</i>			
	duration_mean			
	Poolé	FE (route)	RE	FE (two-way)
	(1)	(2)	(3)	(4)
Congestion (départs/jour)	0.069*** (0.005)	−0.008*** (0.001)	−0.003*** (0.001)	−0.010*** (0.001)
Vols route-jour	−3.658*** (0.181)	0.227*** (0.027)	0.176*** (0.027)	0.229*** (0.027)
Jour : mar	1.050*** (0.243)	−0.534*** (0.073)	−0.507*** (0.073)	
Jour : mer	1.599*** (0.200)	−0.203*** (0.063)	−0.187*** (0.063)	
Jour : jeu	−0.469* (0.256)	0.190*** (0.063)	0.187*** (0.063)	
Jour : ven	−2.679*** (0.325)	0.226*** (0.071)	0.170** (0.071)	
Jour : sam	−1.845*** (0.257)	0.051 (0.064)	0.005 (0.064)	
Jour : dim	−1.280*** (0.287)	−0.038 (0.065)	−0.051 (0.065)	
Mois : juin	0.679*** (0.126)	−0.284*** (0.063)	−0.294*** (0.063)	
Mois : juil	0.525*** (0.081)	−0.143*** (0.042)	−0.142*** (0.042)	
Constant	99.078*** (0.823)		99.465*** (0.558)	
Observations	507,552	507,552	507,552	507,552
R ²	0.035	0.001	0.005	0.0002

Note:

Erreurs standards entre parenthèses, clusterisées par route.

Lecture des résultats :

1. **Modèle poolé (colonne 1)** — Le coefficient associé à la congestion au départ est positif et statistiquement significatif (0.069), suggérant qu’une congestion plus élevée est associée à une durée moyenne plus longue. Toutefois, cette estimation ne contrôle pas l’hétérogénéité non observée entre routes et doit être interprétée avec prudence. Le faible pouvoir explicatif global du modèle (0.035, 0.035) reflète les limites de cette spécification.
2. **Modèle à effets fixes route (colonne 2)** — Une fois l’hétérogénéité invariante dans le temps contrôlée, le coefficient de congestion devient négatif et reste significatif (-0.008). Ce changement de signe suggère que l’association positive observée en OLS reflète principalement des différences structurelles entre routes. La congestion explique une part très limitée de la variation temporelle intra-route de la durée moyenne.
3. **Modèle à effets aléatoires (colonne 3)** — L’estimation fournit un coefficient intermédiaire (-0.003), mais repose sur l’hypothèse forte d’absence de corrélation entre les effets individuels et les régresseurs, hypothèse testée ultérieurement à l’aide d’un test de Hausman.
4. **Modèle à effets fixes bidirectionnels (colonne 4)** — Le coefficient de congestion demeure négatif et significatif (-0.01). Cette spécification, la plus exigeante, contrôle simultanément l’hétérogénéité entre routes et les chocs temporels communs et constitue la référence pour l’interprétation des résultats.

Analyse du résultat contre-intuitif

Le changement de signe du coefficient de congestion entre le modèle poolé et les spécifications à effets fixes suggère l’existence d’un **biais de composition**. Les estimations OLS semblent principalement capter des différences structurelles entre routes, tandis que les modèles à effets fixes identifient un effet conditionnel fondé sur la variation temporelle intra-route.

Le coefficient négatif observé dans les modèles à effets fixes peut refléter des **ajustements opérationnels** ou une **endogénéité résiduelle** de la congestion, celle-ci étant potentiellement corrélée à d’autres facteurs non observés affectant la durée des vols. En conséquence, bien que rigoureux sur le plan technique, le modèle à effets fixes bidirectionnels ne permet pas d’interpréter cet effet comme strictement causal. Les résultats doivent donc être interprétés avec prudence.

4.3 Interprétation économique

4.3.1 Effet marginal

En retenant le modèle à effets fixes bidirectionnels, le coefficient associé à la congestion au départ est de -0.01. Cette estimation implique qu’une augmentation de 10 départs quotidiens à l’aéroport de départ est associée à une variation moyenne de la durée de vol de -0.1 minutes (environ 6 secondes), toutes choses égales par ailleurs.

Ce résultat, de signe négatif, doit être interprété avec prudence. Il ne saurait être compris comme un effet causal direct de la congestion sur la réduction des durées de vol, mais plutôt comme une association conditionnelle identifiée à partir de la variation temporelle intra-route. Il est susceptible de refléter des mécanismes d’endogénéité résiduelle ou des ajustements opérationnels non observés.

4.3.2 Limites de l’interprétation causale

Compte tenu du résultat contre-intuitif, nous devons reconnaître les limites importantes de notre analyse :

1. **Absence d’instrument** : Sans variable instrumentale pour la congestion, nous ne pouvons pas établir de lien causal crédible.
2. **Mesure composite de la durée** : La variable `duration_mean` agrège plusieurs phases qui peuvent réagir différemment à la congestion.

3. **Endogénéité temporelle** : La congestion quotidienne est corrélée avec des facteurs non observés (météo, événements, optimisation) qui affectent aussi les durées.

4. **Nécessité d'analyses complémentaires** :

- Décomposition de la durée par phase de vol
- Variables instrumentales (ex : capacité réglementaire, fermetures de pistes planifiées)
- Analyse de discontinuité (seuils de saturation)
- Étude qualitative des procédures opérationnelles

4.4 Tests de spécification

Table 6: Tests de spécification économétrique

Test	Statistique	Valeur-p	Conclusion
Breusch-Pagan (Poolé vs RE)	12564126.59	0	Rejeter poolé, préférer RE
Hausman (FE vs RE)	674.35	0	Rejeter RE, préférer FE

Interprétation :

1. Le test de Breusch-Pagan rejette l'hypothèse nulle d'absence d'effets individuels (statistique = 1.2564127×10^7 , valeur-p < 0,01), indiquant que le modèle poolé est inadapté et qu'un modèle avec effets individuels est requis.

2. Par ailleurs, le test de Hausman rejette l'hypothèse d'exogénéité des effets individuels (statistique = 674.35, valeur-p < 0,01), ce qui conduit à privilégier une spécification à effets fixes plutôt qu'à effets aléatoires.

Conclusion des tests : Ces résultats confortent le choix du modèle à effets fixes, et en particulier de la spécification à effets fixes bidirectionnels retenue pour l'analyse principale.

5 Choix du modèle préféré et justification

5.1 Arbitrage entre les spécifications

5.1.1 Critères de sélection

Le choix du modèle préféré repose sur plusieurs critères complémentaires :

1. **Validité interne (identification causale)** : Capacité à contrôler les biais d'endogénéité et d'hétérogénéité non observée.
2. **Tests statistiques formels** : Résultats des tests de Breusch-Pagan et Hausman.
3. **Robustesse théorique** : Cohérence avec la littérature économique et les mécanismes attendus.
4. **Efficacité statistique** : Précision des estimations (erreurs standards).

5.1.2 Analyse comparative

Table 7: Comparaison des spécifications économétriques

Modèle	Contrôle hétérogénéité route	Contrôle chocs temporels	Test BP/Hausman	Coefficient congestion
Poolé	Non	Partiel	Rejeté	0.069***
FE (route)	Oui	Partiel	Préféré	-0.008***
RE	Partiel	Partiel	Rejeté	-0.003***
FE (two-way)	Oui	Oui	Préféré	-0.010***

5.2 Modèle préféré : FE two-way (route + date)

Nous retenons le **modèle à effets fixes bidirectionnels (route et date)** comme spécification privilégiée d'un point de vue méthodologique. En contrôlant simultanément l'hétérogénéité invariante entre routes et les chocs temporels communs, cette approche constitue la spécification la plus exigeante et conforme aux pratiques standards de l'économétrie de panel.

Les tests de spécification (Breusch–Pagan et Hausman) conduisent à rejeter le modèle poolé et le modèle à effets aléatoires au profit des effets fixes. Le coefficient associé à la congestion reste statistiquement significatif dans toutes les spécifications à effets fixes, y compris dans le modèle two-way, ce qui atteste de la robustesse statistique de l'association estimée.

Toutefois, le signe négatif du coefficient doit être interprété avec prudence. Il reflète une **association conditionnelle**, identifiée à partir de la variation temporelle intra-route, et non un effet causal direct de la congestion sur la durée des vols. La congestion quotidienne demeure potentiellement endogène et corrélée à des facteurs non observés (conditions météorologiques, organisation opérationnelle), qui ne sont pas entièrement capturés par les effets fixes.

En conséquence, le modèle FE two-way doit être interprété comme le **meilleur modèle descriptif disponible**, offrant un compromis robuste entre contrôle des biais et exploitation de la variation temporelle, mais ne permettant pas d'établir une relation causale stricte sans recours à des stratégies d'identification supplémentaires.

5.3 Tests de robustesse

5.3.1 Spécification logarithmique

Pour tester la sensibilité de nos résultats à la forme fonctionnelle, nous estimons une variante avec transformation logarithmique de la congestion :

$$Duration_{it} = \alpha_i + \delta_t + \beta_1 \log(1 + Congestion_{it}) + \beta_2 NbFlights_{it} + \mathbf{X}'_{it}\gamma + \varepsilon_{it}$$

Afin de tester la sensibilité des résultats à la forme fonctionnelle, nous estimons une spécification alternative dans laquelle la congestion au départ est introduite sous forme logarithmique. Dans cette spécification, le coefficient associé à $\log(1 + Congestion)$ est estimé à 1.134 et est statistiquement significatif ($p = 0$).

Ce résultat suggère qu'une augmentation proportionnelle de la congestion est associée à une variation de la durée moyenne des vols en niveau. Toutefois, le changement de signe par rapport aux spécifications linéaires indique que l'effet estimé est sensible à la forme fonctionnelle retenue. Cette sensibilité renforce l'interprétation prudente des résultats principaux et confirme que les estimations doivent être comprises comme des associations conditionnelles plutôt que comme des effets causaux stricts.

5.3.2 Spécification par quintiles

Pour évaluer la non-linéarité de l'effet, nous remplaçons la variable continue de congestion par des indicatrices de quintiles :

$$Duration_{it} = \alpha_i + \delta_t + \sum_{q=2}^5 \theta_q \mathbb{1}(Congestion_{it} \in Q_q) + \beta_2 NbFlights_{it} + \mathbf{X}'_{it}\gamma + \varepsilon_{it}$$

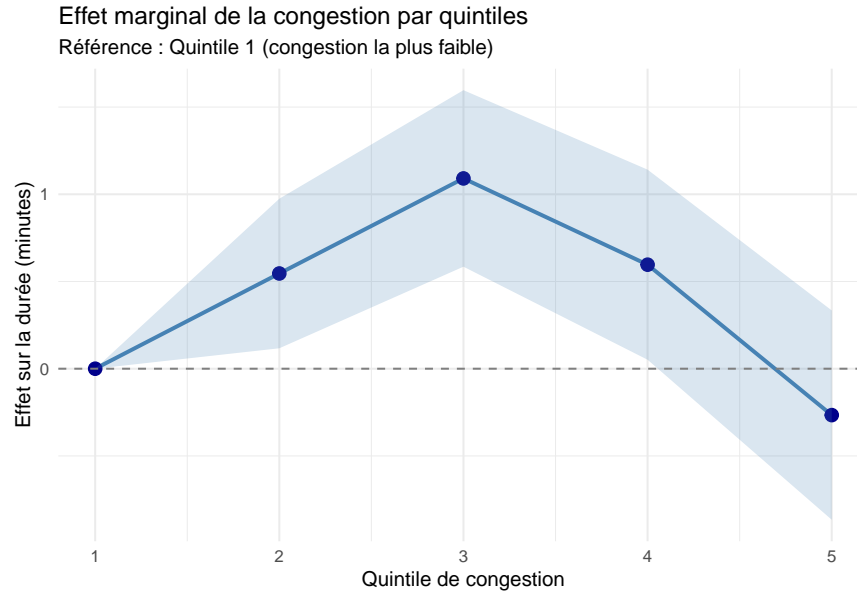


Figure 4: Effet marginal de la congestion par quintiles (FE route)

Interprétation : L'estimation par quintiles met en évidence une **relation non linéaire** entre congestion et durée moyenne des vols. L'effet estimé est positif et croissant entre les premiers quintiles, atteignant un maximum pour des niveaux intermédiaires de congestion, puis s'atténue aux niveaux les plus élevés. L'effet associé au quintile le plus congestionné est plus faible et statistiquement moins précis, suggérant que l'impact marginal de la congestion n'est pas monotone. Ces résultats confirment l'existence d'une non-linéarité, tout en appelant à une interprétation prudente des effets aux niveaux extrêmes de congestion.

5.4 Conclusion sur le choix de spécification

L'ensemble des analyses (tests de spécification, tests de robustesse et analyse de non-linéarité) conduit aux conclusions suivantes. Le modèle à effets fixes bidirectionnels (FE two-way) constitue la spécification la plus exigeante sur le plan méthodologique, en contrôlant simultanément l'hétérogénéité invariante entre routes et les chocs temporels communs.

Les résultats indiquent que l'association entre congestion et durée moyenne des vols est sensible à la forme fonctionnelle retenue et présente une non-linéarité, l'effet estimé étant principalement concentré aux niveaux intermédiaires de congestion. Ces éléments confirment que les estimations doivent être interprétées comme des effets conditionnels, et non comme des effets causaux stricts.

D'un point de vue économique, ces résultats suggèrent que les politiques de gestion de la congestion doivent tenir compte de l'hétérogénéité des situations opérationnelles, en particulier de la complexité des mécanismes à l'œuvre dans les aéroports fortement congestionnés.

6 Conclusion générale

Cette étude analyse l’association entre congestion aéroportuaire et durée effective des vols en Europe à partir de données opérationnelles d’Eurocontrol sur la période mai–juillet 2023. L’exploitation d’un panel route–jour et de différentes spécifications économétriques met en évidence un résultat central.

Les modèles à effets fixes, en particulier la spécification bidirectionnelle, conduisent à un coefficient négatif associé à la congestion, estimé à -0.01. Ce résultat, opposé à celui obtenu en modèle poolé, souligne l’importance des effets de composition et de l’hétérogénéité structurelle entre routes. Il suggère que les estimations naïves capturent principalement des différences persistantes entre aéroports et types de liaisons.

Toutefois, l’analyse met en évidence des limites importantes d’interprétation causale. La congestion quotidienne apparaît potentiellement endogène et corrélée à des facteurs non observés affectant simultanément la durée des vols. Les tests de robustesse montrent en outre une sensibilité des résultats à la forme fonctionnelle et une non-linéarité de l’effet estimé.

Dans ce contexte, le modèle à effets fixes bidirectionnels constitue la spécification la plus rigoureuse pour documenter une association conditionnelle, sans permettre d’identifier un effet causal strict. Ces résultats soulignent la nécessité de stratégies d’identification complémentaires et de données plus fines pour éclairer l’analyse économique de la congestion aéroportuaire.

7 Annexes

7.1 Annexe A : Code R complet

Note : Le code complet est disponible dans le fichier source .Rmd joints au dossier.

7.2 Annexe B : Statistiques descriptives complémentaires

Table 8: Distribution du nombre de jours d’observation par route

Jours min	Jours médiane	Jours moyenne	Jours max
20	50	54.5	92

Interprétation : Le panel est déséquilibré, avec une forte hétérogénéité du nombre de jours par route. Certaines routes (liaisons saisonnières ou charter) ne sont opérées que 20-25 jours, tandis que les liaisons régulières des hubs majeurs sont observées quotidiennement (92 jours).

Fin du document