



Université Paris-Est Créteil  
Master 1 Économie appliquée parcours MASERATI

# **Existe-t-il des asymétries dans les ajustements des prix à la pompe ?**

Auteur du Mémoire :  
**Gaël PEFOURQUE**

Directeur de Mémoire :  
**Ferhat MIHOUBI**

**2023 – 2024**

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Revue de littérature</b>	<b>3</b>
2.1	Résultats empiriques et hypothèses sur les causes des asymétries . . . . .	3
2.2	Études sur le marché français . . . . .	4
2.3	Approche à l'échelle des stations-service . . . . .	5
<b>3</b>	<b>Étude empirique</b>	<b>6</b>
3.1	Pétrole, carburants et fiscalité . . . . .	6
3.1.1	Production du pétrole brut et marché des carburants . . . . .	6
3.1.2	La fiscalité des carburants en France . . . . .	7
3.2	Présentation des données . . . . .	7
3.2.1	Base de données . . . . .	7
3.2.2	Statistiques descriptives . . . . .	8
3.3	Méthode utilisée . . . . .	15
<b>4</b>	<b>Résultats et discussion</b>	<b>17</b>
4.1	Résultats . . . . .	17
4.2	Discussion . . . . .	23
4.3	Limites . . . . .	23
4.3.1	Critère de compétition . . . . .	23
4.3.2	Explications de l'asymétrie . . . . .	24
<b>5</b>	<b>Conclusion</b>	<b>24</b>
<b>Bibliographie</b>		<b>25</b>
<b>Annexes</b>		<b>26</b>
5.1	. . . . .	26
5.2	. . . . .	27
<b>Code Python</b>		<b>28</b>

# 1 Introduction

Le 8 février 2023, TotalEnergies annonçait un bénéfice record de 20,5 milliards de dollars au titre de l'exercice 2022. La question de la taxation des superprofits pour les entreprises ayant profité de la hausse des prix de l'énergie et de l'inflation, s'est imposée dans le débat public français. Outre-Atlantique également, la société pétrolière ExxonMobil a dégagé un profit exceptionnel en 2022 de 59 milliards de dollars. En réaction, Joe Biden a dénoncé ces bénéfices record par "Big Oil" (ExxonMobil, Chevron, BP, Shell et TotalEnergies), et a proposé « de quadrupler la taxe sur les rachats d'actions des sociétés [actuellement de 1 %] afin d'encourager plutôt les investissements à long terme ». ([Leparmetier \(2023\)](#)).

Ces profits sont la conséquence de la hausse des prix du pétrole brut. En effet, en 2022, le cours du baril de Brent, ainsi que l'ensemble des produits pétroliers, ont connu une forte augmentation, en raison de la reprise économique post-COVID-19, couplée aux tensions géopolitiques de la guerre en Ukraine. Le prix moyen du baril sur l'année 2022 a augmenté de 43 % par rapport à 2021, passant à 101 dollars du baril contre 70,62 dollars l'année précédente ([Misak \(2023\)](#)). Cette hausse s'explique par plusieurs facteurs : la reprise du trafic aérien suite à la fin de l'épidémie de COVID-19 et plus globalement l'augmentation de la demande mondiale, les tensions géopolitiques liées à l'invasion en Ukraine et l'embargo sur le pétrole et le gaz russes, ainsi que la politique de limitation de l'offre de pétrole des pays de l'OPEP (Organisation des pays exportateurs de pétrole) ont tiré vers le haut le Brent et l'ensemble des produits pétroliers, atteignant des niveaux records.

Dans ce contexte, le secteur pétrolier est accusé d'augmenter rapidement les prix "à la pompe" suite à la hausse des cours du pétrole brut, et de les diminuer lentement à la suite d'une baisse des cours du brut. Cette accusation n'est pas nouvelle, elle revient à chaque crise et souligne l'intérêt marqué de l'opinion publique et des responsables politiques pour l'industrie pétrolière et les profits élevés qu'elle génère. La manière dont les variations positives et négatives du cours du pétrole brut se répercutent sur les prix des carburants est ainsi une question actuelle et essentielle pour de nombreux consommateurs, pour lesquels le carburant représente une dépense significative.

Le but de ce mémoire est donc de répondre à la problématique suivante : existe-t-il des asymétries dans les ajustements des prix à la pompe ? Le caractère actuel de cette question, ainsi que les implications politiques et économiques concrètes de l'existence d'asymétrie dans l'ajustement des prix de détail des carburants – plafonnement des prix, contrôle des marges, taxation des superprofits – démontrent l'intérêt d'apporter des réponses à cette question.

Le reste du mémoire est organisé comme suit : la section 2 dresse un état des lieux de la littérature économique relative à ce sujet. La revue de littérature permet de synthétiser les résultats des recherches précédentes, et d'analyser les différentes méthodologies employées, ainsi que leurs implications. La section 3 traite des différentes notions liées au pétrole et aux carburants. Ensuite, elle expose la base de données et la méthode utilisées dans ce mémoire pour étudier la présence d'asymétrie dans les variations des prix à la pompe. La section 4 présente les résultats et discutera des implications pratiques et théoriques qui en découlent. Enfin, la section 5 conclut ce mémoire.

## 2 Revue de littérature

### 2.1 Résultats empiriques et hypothèses sur les causes des asymétries

L'idée que les prix des stations-service augmentent rapidement mais diminuent lentement semble être communément admise parmi les consommateurs. En réponse à ce sentiment, une vaste littérature à propos de la relation entre les prix des carburants et les cours du pétrole a tenté d'établir la présence d'asymétrie, à partir de données diverses et de modèles économétriques variés. [Bacon \(1991\)](#) a été le premier à étudier ce phénomène et a inventé l'expression "rockets and feathers" pour décrire cette situation : les prix décolleraient comme des fusées et retomberaient comme des plumes. L'étude de [Bacon \(1991\)](#) porte sur le marché britannique des stations-service. A l'aide de données bimensuelles sur la période 1987–1989, il analyse la réponse des prix du diesel face aux variations du cours du Brent. Il trouve que les prix du diesel s'ajustent plus rapidement quand le Brent augmente que lorsqu'il diminue, et montre donc l'existence d'asymétrie sur ce marché, confirmant ainsi la croyance commune à ce sujet. Depuis cette étude, beaucoup de travaux se sont intéressés à cette question.

[Manning \(1991\)](#) étudie aussi le cas britannique à l'aide d'un modèle à correction d'erreur (ECM) et conclut aussi à la présence d'asymétrie avec des données mensuelles sur la période 1973–1988. Pour le cas américain, [Karrenbrock \(1991\)](#) analyse le cas américain et utilise des données mensuelles pour la période 1983–1990. Il ne trouve pas d'asymétrie avec un modèle ARDL, et trouve même que les prix s'ajustent plus rapidement à la baisse qu'à la hausse. [Shin \(1994\)](#) étudie à partir des prix agrégés mensuels aux Etats-Unis pour la période 1982–1990 et ne trouve lui aussi aucune asymétrie avec un modèle dynamique, que ce soit pour l'étape pétrole brut - raffinerie, ou l'étape raffinerie - stations-service dans la chaîne d'approvisionnement.

[Borenstein et al. \(1997\)](#) s'intéressent au cas du diesel sur le marché américain avec des données hebdomadaires pour la période 1988–1992. Ils concluent à la présence d'asymétrie utilisant un modèle à correction d'erreur : les prix du diesel réagissent en 4 semaines à une hausse du brut contre 8 semaines pour une baisse du brut. Leur étude montre que les asymétries n'apparaissent pas au niveau du raffinage, mais au niveau des stations-services.

Les auteurs fournissent trois interprétations pour expliquer l'asymétrie de réponse du diesel sur le marché américain. La première hypothèse est d'asymétrie du coût d'ajustement des stocks face à une hausse et une baisse de la demande (déjà évoquée par [Reagan and Weitzman \(1982\)](#)). Lors d'une hausse imprévue de la demande, les firmes ne peuvent pas toujours réduire leurs stocks d'un montant égal à l'accroissement de la demande (stocks non négatifs). Ainsi, le prix augmente rapidement pour freiner l'augmentation de la demande. En revanche, face à une baisse imprévue de la demande, les firmes n'ont pas de coût d'approvisionnement car elles n'ont pas besoin d'augmenter leurs stocks. Elles diminuent donc faiblement les prix, ce qui expliquerait l'asymétrie des prix à la pompe, moins sensibles à une baisse de la demande (soit une hausse du cours du pétrole brut).

[Borenstein et al. \(1997\)](#) fournissent une deuxième explication possible. Les firmes ne possèdent qu'une information limitée et imparfaite à propos des prix de leurs concurrents. Quand les prix du brut diminuent, les firmes font attention au signal de prix qu'elles enverraient à leurs concurrents en abaissant leurs prix, pour ne pas donner l'impression de mener une guerre commerciale pour gagner en part de marché (voir aussi [Lewis \(2011\)](#)). Les firmes augmentent rapidement leurs prix en cas de hausse du brut pour maintenir leurs marges, mais face à une baisse du prix du brut, elles adoptent des considérations stratégiques, pour ne pas engendrer une guerre des prix, qui retardent donc l'ajustement des prix des carburants à la baisse. Dans cette hypothèse, le comportement de collusion et la situation d'oligopole du marché sont donc la source de l'asymétrie.

La troisième hypothèse avancée par [Borenstein et al. \(1997\)](#) et [Johnson \(2002\)](#), est que la volatilité

des cours du pétrole brut ainsi que le nombre de station-services brouillent la visibilité des agents économiques. Les consommateurs font face à des coûts d'information et de recherche qui affectent leurs choix. Ainsi, quand les prix augmentent, les agents sont attentifs aux prix des différentes stations-services et cherchent à minimiser leurs coûts. En cas de baisse des prix, les agents recherchent moins intensément les meilleurs prix. Prenons le cas d'un conducteur : il passe devant une station-service et constate que le prix du diesel est plus bas qu'il ne le pensait. Il s'arrête faire le plein plus facilement à la première station qu'il rencontre, que si le prix du diesel avait augmenté. Dans cette hypothèse, l'asymétrie est du côté de la demande, et le changement dans l'attention portée par les consommateurs aux "bonnes affaires" en cas de baisse des prix explique l'asymétrie des prix du diesel. [Tappata \(2009\)](#), [Yang and Ye \(2008\)](#) et [Lewis \(2011\)](#) développent des modèles dans lesquels les coûts de recherche causent l'asymétrie des prix du diesel. Dans cette hypothèse, l'asymétrie d'ajustement ne vient pas du côté des producteurs mais des consommateurs, et n'est pas liée ni au pouvoir de marché des stations-service, ni à une situation oligopolistique comme dans l'hypothèse précédente.

La publication de [Borenstein et al. \(1997\)](#) a grandement influencé les études ultérieures sur ce sujet, qui ne se sont plus contentées de constater la présence ou l'absence d'asymétrie des prix des carburants (principalement du diesel), mais ont tenté de comprendre la source de l'asymétrie, lorsqu'observée. La littérature étant importante à ce sujet et les méthodologies variées, plusieurs articles ont effectué une revue de la littérature existante ([Frey and Manera \(2007\)](#), [Grasso and Manera \(2007\)](#), [Perdiguero-García \(2013\)](#), [Ederington et al. \(2019\)](#)) pour tenter de rendre compte des conclusions les plus communément observées.

Dans leur étude, [Frey and Manera \(2007\)](#) montrent que sur 69 articles et 83 modèles estimés, seulement 11 modèles ne montrent aucune preuve d'asymétrie. Le consensus de la littérature est donc la présence d'écart significatifs de réponse des prix des carburants aux variations des cours du brut. Les modèles les plus employés sont les modèles à correction d'erreur – ECM – (36 modèles), les modèles autorégressifs à retards échelonnés ou distribués – ARDL – (21), et les modèles à vecteur autorégressif – VAR – (8). Les trois pays les plus étudiés sont les États-Unis, le Royaume-Uni et l'Allemagne. La quasi-totalité des études portent sur des pays occidentaux, avec seulement six études dédiées à des pays non-occidentaux.

[Perdiguero-García \(2013\)](#) examine 403 différents résultats et analyse les caractéristiques de chaque étude pour tester si les résultats dépendent des modèles estimés et des types de données utilisées. Environ 60 % des résultats étudiés rejettent l'hypothèse de symétrie. Les conclusions de cette étude sont diverses. [Perdiguero-García \(2013\)](#) trouve que les données mensuelles tendent à masquer la présence d'asymétrie dans les données, de même pour les données contenant des prix agrégés. Les données avec une granularité plus fine et des prix au niveau microéconomique rendent mieux compte des asymétries. Ce résultat semble logique, et en accord avec les conclusions d'autres études (voir [Bachmeier and Griffin \(2003\)](#), [Bettendorf et al. \(2003\)](#), [Kilian \(2010\)](#)). Par ailleurs, [Perdiguero-García \(2013\)](#) explique que les études sur des périodes récentes sont associées à une plus faible probabilité d'observation d'asymétrie des prix à la pompe, et associe cela à la libéralisation du marché et à des niveaux de compétition plus élevés. Dans son étude, [Perdiguero-García \(2013\)](#) ne contrôle pas le cours du pétrole brut de référence utilisé, tel que le WTI, le Brent ou le Dubai, qui aurait constitué une variable pertinente. En effet, il aurait été intéressant d'évaluer si le choix d'un pétrole brut de référence influait statistiquement sur l'observation d'asymétrie, à l'échelle de la littérature.

## 2.2 Études sur le marché français

La très grande majorité des travaux portent sur le marché américain (26 sur 34 dans [Frey and Manera \(2007\)](#)). L'analyse du marché français présente un intérêt car les marchés nationaux diffèrent sur différents aspects – notamment leur structure ou leur niveau de libéralisation – et cela peut impacter

la présence d'asymétrie. A ma connaissance, il existe cinq études traitant de l'asymétrie des prix des carburants sur le cas français spécifiquement : [Audenis et al. \(2002\)](#), [Lamotte et al. \(2013\)](#), [Gautier and Saout \(2015\)](#), [Boroumand et al. \(2016\)](#), et [Boroumand et al. \(2021\)](#).

[Audenis et al. \(2002\)](#) utilisent des données agrégées mensuelles pour la période 1980–2000, pour le SP95 (sans-plomb 95), diesel et fioul domestique. Les auteurs estiment un modèle à correction d'erreur proche de [Borenstein et al. \(1997\)](#), et montrent que les prix s'ajustent plus rapidement à la hausse qu'à la baisse, engendrant ainsi un coût supplémentaire pour les consommateurs pendant plusieurs mois. Ils expliquent cela par la caractère oligopolistique du marché.

[Lamotte et al. \(2013\)](#) étudient l'ajustement du diesel et du SP95 sur le marché français avec des données hebdomadaires pour la période 1990–2011. Ils trouvent en estimant un modèle ARDL de l'asymétrie dans les deux types de carburant. Les variations du Brent ont des effets plus prolongés sur les prix du SP95 que sur les prix du diesel.

[Gautier and Saout \(2015\)](#) utilisent des données microéconomiques provenant de relevés de prix quotidiens des stations-service françaises pour la période 2007–2009. Ils ne concluent pas à la présence d'asymétrie dans la transmission du brut vers le diesel ou le SP95, et estiment qu'un choc est intégralement transmis aux prix des carburants en 10 jours. Ce résultat est étonnant et n'est pas en phase avec les conclusions les plus courantes dans la littérature.

[Boroumand et al. \(2016\)](#) estiment un modèle de Markov (Markov-switching model) avec des données hebdomadaires pour la période 1990–2011. Ils trouvent bien de l'asymétrie entre les prix des carburants et du brut, et ajoutent qu'en période de faible volatilité, les prix des carburants répondent plus aux variations en amont qu'en période de haute volatilité.

[Boroumand et al. \(2021\)](#) utilisent des données hebdomadaires pour la période 2005–2020 pour les prix du diesel et du SP95. Ils estiment un modèle ARDL mais se concentrent sur 3 années avec des chutes importantes des cours du pétrole brut : 2008 (crise financière), 2014 (chute du baril en raison de l'exploitation du gaz de schiste américain), et 2020 (crise du COVID-19). Les auteurs observent de l'asymétrie entre les variations du Brent et les prix du diesel et du SP95 pour toute la période, mais montrent un pic d'asymétrie lors de la crise financière de 2008.

4 des 5 études indiquent la présence d'asymétrie dans le cas français. Elles couvrent la période 1985–2020. La seule étude utilisant des données microéconomiques ([Gautier and Saout \(2015\)](#)) s'arrête en 2009 et ne trouve pas d'asymétrie dans les prix à la pompe. Il serait donc intéressant d'analyser à nouveau la présence d'asymétrie en France avec des données microéconomiques actuelles, pour évaluer la validité des conclusions de [Gautier and Saout \(2015\)](#).

## 2.3 Approche à l'échelle des stations-service

La plupart des études à propos de l'asymétrie de réponse des prix des carburants aux variations des cours du pétrole brut utilisent des données agrégées ([Ederington et al. \(2019\)](#)). Comme l'a montré [Perdiguero-García \(2013\)](#), des données agrégées ont tendance à ne pas trouver autant d'asymétrie que des données au niveau microéconomique. L'utilisation majoritaire de ce type de données s'explique par la complexité technique de recueillir des données provenant directement de stations-service. De plus, les données à l'échelle des station-services permettent de contrôler des variables liées au type de station, à la concurrence, à la localisation. Ces variables sont importantes et jouent un rôle sur la présence ou non d'asymétrie, selon la théorie économique. Plusieurs études ont donc utilisé des données microéconomiques dans le but de mieux expliquer les causes de l'asymétrie, plutôt que de s'arrêter simplement à son observation.

[Balmaceda and Soruco \(2008\)](#) étudie le cas du marché de détail du diesel dans la ville de Santiago, au Chili. L'auteur utilise les relevés de prix hebdomadaires pour 44 stations-service sur la période 2001–2004. Il estime un modèle à correction d'erreur classique, et contrôle la localisation et la marque de chaque station-service. [Balmaceda and Soruco \(2008\)](#) trouve bien une asymétrie de réponse des prix des carburants aux variations des prix du pétrole brut. Aussi, il montre que les stations "de marque" (par exemple Shell, Cope) ont un taux de marge supérieur aux stations "sans marque". Pour autant, les stations avec une forte marge n'ont pas des comportements de prix plus asymétriques que les stations qui ont une faible marge, contrairement à ce qu'avait trouvé [Deltas \(2008\)](#) pour le marché américain.

[Remer \(2015\)](#) analyse les prix quotidiens de 11 000 stations-service dans plusieurs états américains (Maryland, New Jersey, Philadelphia, Virginia, Washington, et Washington DC) pour la période 2008–2009 et constate de l'asymétrie. Il trouve aussi que les stations-service indépendantes pratiquent de prix plus bas que les stations-service "de marque", ce qui concorde avec les résultats de [Balmaceda and Soruco \(2008\)](#) et [Hosken et al. \(2008\)](#). De plus, il confirme que le nombre de stations concurrentes aux alentours fait baisser les prix pratiqués, et que le prix augmente avec la distance de la station concurrente la plus proche, ce qui confirme l'hypothèse de pouvoir de marché des stations-service. [Remer \(2015\)](#) montre aussi que les prix des supercarburants prennent plus de temps à diminuer suite à une baisse du brut que les prix du diesel, car le marché des supercarburants est moins concurrentiel dans son échantillon. Cela contraste avec les conclusions de [Liu et al. \(2010\)](#) qui trouvent que le marché du diesel est moins concurrentiel que celui des supercarburants.

[Balaguer and Ripollés \(2016\)](#) étudient les ajustements des prix d'un panel de stations-service en Espagne pour la période 2010–2012. Ils analysent deux aires urbaines de respectivement 283 et 185 stations-service à l'aide des relevés de prix quotidiens. Ils constatent bien de l'asymétrie pendant une semaine après un choc à l'échelle microéconomique. Pourtant lorsqu'ils agrègent les données, les auteurs ne peuvent pas rejeter l'hypothèse de symétrie. Cela peut contribuer à expliquer que la présence d'asymétrie n'ait pas toujours été observée par des études travaillant à partir de données agrégées au niveau national ou régional, et souligne l'intérêt de l'usage de données à l'échelle des stations-service.

La revue de littérature met en évidence qu'il existe peu d'études sur le cas français des asymétries des prix des carburants face aux variations des prix du pétrole brut. Par ailleurs, la plupart des travaux de la littérature utilisent des données au niveau macroéconomique avec une fréquence modérée (hebdomadaire ou mensuelle), ce qui a des implications sur la détection d'asymétries dans certains cas (voir [Perdiguero-García \(2013\)](#) et [Balaguer and Ripollés \(2016\)](#)). Ce mémoire a pour but de contribuer à combler une lacune dans la littérature récente en utilisant des données de haute fréquence à l'échelle des stations-service pour le cas français. La dernière étude de ce type sur le marché français ([Gautier and Saout \(2015\)](#)) utilise des données antérieures à 2010 et conclut à la non-présence d'asymétrie, ce qui n'est pas en accord avec les autres travaux sur le marché français. Il est donc pertinent de se pencher sur cette question pour enrichir la littérature économique.

### 3 Étude empirique

#### 3.1 Pétrole, carburants et fiscalité

##### 3.1.1 Production du pétrole brut et marché des carburants

Le pétrole brut est le principal composant des produits pétroliers, notamment des carburants. Il est extrait de l'exploitation de puits de pétrole et est un mélange d'hydrocarbures. Il existe trois bruts

de référence qui servent de points de repère pour le prix des autres pétroles bruts sur les marchés mondiaux : le Brent (exploité en mer du Nord) ; le WTI (exploité aux États-Unis) ; le Dubaï Light (exploité en Asie). Le baril de Brent fixe un prix standard pour 78 % des gisements de pétrole brut mondial et sert de référence en Europe, au Moyen-Orient et en Afrique. C'est donc le cours de référence que nous allons utiliser dans ce mémoire.

Le pétrole brut est ensuite raffiné pour pouvoir servir dans les moteurs à combustion (étape de production). A la suite de ce traitement, le pétrole est transformé en gazole ou en essence, selon le processus de raffinage utilisé. Les carburants sont ensuite transportés, stockés et distribués vers les stations-service qui les vendent ensuite (étape de distribution). Chacune de ces phases entre l'extraction du pétrole brut et la vente au détail par les stations-service contribue à expliquer l'écart entre pétrole brut et carburants. Toutes les étapes sont donc des sources potentielles d'asymétrie dans les variations des prix des carburants face aux variations du brut.

### 3.1.2 La fiscalité des carburants en France

Les prix des carburants contiennent ainsi le coût du pétrole brut, les coûts de production, la marge du distributeur, et les taxes auxquelles sont soumis les différents carburants. Dans le cas du marché français, les taxes comptent pour 60 % du prix final du gazole et de l'essence (Ministère de la transition écologique). Les deux principales taxes sont :

- La taxe intérieure de consommation sur les produits énergétiques (TICPE, anciennement TIPP) : elle est la principale taxe sur les carburants en France. C'est un droit d'accise et s'élève en 2022 pour l'Île-de-France à 62,64 c€/l pour le gazole ; 70,04 c€/l pour le SP95 et le SP98 ; et 68,04 c€/l pour le SP95-E10. La TICPE représente approximativement 40 % du prix final.
- La taxe sur la valeur ajoutée (TVA) : d'un taux de 20 %, la taxe s'applique sur le prix HT, auquel s'ajoute la TICPE. La TVA représente approximativement 20 % du prix final.

## 3.2 Présentation des données

### 3.2.1 Base de données

La base de données que l'on va utiliser contient les prix individuels de toutes les stations-service vendant plus de 500m<sup>3</sup> par an de carburants en France. Les stations-service ont l'obligation légale depuis le 1<sup>er</sup> janvier 2007 d'afficher toute modification de prix du gazole (diesel), SP95-E5 (SP95), SP95-E10, SP98, E85 et GPLc sur le site internet gouvernemental [www.prix-carburants.gouv.fr](http://www.prix-carburants.gouv.fr). Ces informations sont accessibles en Open Data de manière libre et gratuite. Les principales variables dans cette base de données sont : le prix de vente en euros par litre incluant les taxes (TVA et TICPE), le type de carburant, la date de mise à jour du prix, et des informations sur chaque station-service : numéro d'identification, code postal, ville, coordonnées GPS, type de station-service (autoroute ou route). La base de données est constituée des informations pour l'année 2022, soit du 1<sup>er</sup> janvier 2022 au 31 décembre 2022.

Nous récupérons les données pour 2022 du cours du Brent Spot Price FOB d'après les chiffres de l'U.S. Energy Information Administration en dollars. On convertit ensuite le cours du Brent en euros à l'aide du taux de change quotidien USD/EUR d'après les chiffres de Boursorama. Nous divisions enfin le prix du baril en euros par le nombre de litres dans un baril (158,98) pour trouver le prix au litre du brut en euros pour la période 2022. De ce fait, les variations du Brent ne sont pas causées par la fluctuation du taux de change entre le dollar et l'euro car le prix du Brent prend en compte le taux de change et est exprimé en euros.

On calcule aussi les prix à la pompe hors taxe (HT) en retirant la TVA et TICPE propre à chaque carburant et chaque département de la manière suivante :  $P_{HT} = \frac{P_{TTC}}{1+TVA} - TICPE_{locale}$ . Le passage au prix hors taxe permet de ne pas avoir d'effet indésirable lié à un changement de taxation, qui impacterait le prix des carburants. Ici, la fiscalité est restée stable en 2022, mais l'ajustement des prix de l'entreprise étant le résultat de la maximisation de ses profits calculés à partir des prix HT, il est pertinent de raisonner avec les prix HT pour cette étude. Enfin, on élimine les valeurs manquantes et erreurs de saisie de prix, et on calcule pour chaque station-service le nombre de concurrents dans un rayon de 5 kilomètres à l'aide des coordonnées géographiques de la base de données.<sup>1</sup> Cela va permettre de prendre en compte la compétition géographique dans l'analyse de l'asymétrie des prix.

Après traitement des valeurs manquantes et des valeurs aberrantes, l'échantillon pour 2022 contient 9470 stations-service, et 4 537 449 relevés de prix, pour le gazole, SP95, SP95-E10, SP98, E85 et GPLc. L'E85 (biocarburant composé à 85 % d'éthanol) et GPLc (mélange de propane et de butane) étant des carburants alternatifs peu utilisés et faiblement corrélés au prix du baril de pétrole (du fait de leur composition), ne sont pas le centre des analyses de ce mémoire. Les carburants "principaux" sont ainsi le gazole (73 % du volume vendu en 2022), le SP95-E10 (15 %), le SP98 (5 %), et le SP95 (5 %) (d'après l'étude 2022 sur les réseaux de distribution des carburants routiers en France – Ufip Énergies et Mobilités).

Il existe quatre types de stations-service en France : les stations-service de supermarchés (qui représentent 40 % des stations et 60 % du volume vendu)<sup>2</sup>; les stations-service affiliées à des compagnies pétrolières ; les stations-service dites indépendantes ; les stations-services implantées sur autoroute (souvent affiliées à des compagnies pétrolières). Notre base de données ne permet de distinguer que les stations sur route (majoritairement proches de supermarchés), et les stations sur autoroute (majoritairement sous franchise). Cela ne nous permet donc pas de contrôler directement les comportements des enseignes indépendantes face aux stations dites affiliées.

### 3.2.2 Statistiques descriptives

La base de données contient les relevés de prix pour le gazole, SP95-E10, SP95, SP98, E85 et GPLc pour l'année 2022. La figure 1 est un graphique représentant les prix au litre quotidiens moyens hors taxe pour chaque carburant pour 2022, ainsi que le prix au litre du brut (représenté par la courbe en noir). On observe que les prix HT des carburants principaux sont supérieurs au prix du brut, ce qui est cohérent car les stations-service font face à des coûts de distribution, de transport, de raffinage qui augmentent le coût par rapport au cours du Brent. Par ailleurs, on remarque que les courbes du E85 et GPLc (respectivement jaune et verte) semblent décorrélées du cours du Brent, ce qui est cohérent car ces carburants ne sont pas majoritairement composés de pétrole brut.

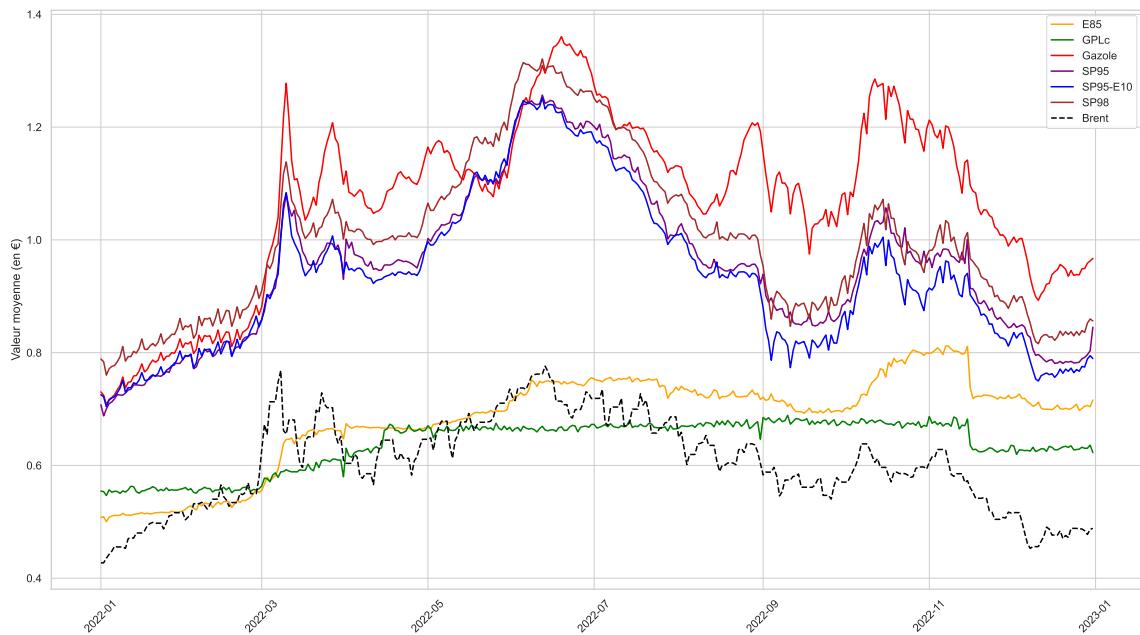
Les prix du pétrole ont été volatiles pour la période 2022, en raison notamment de la guerre en Ukraine et de la reprise économique post-Covid ([Misak \(2023\)](#)). Ces fluctuations apparaissent clairement sur ce graphique, et on observe des pics en mars, juin et octobre. Les mouvements particulièrement importants sur cette période du Brent et des carburants sont intéressants pour notre étude à plusieurs égards. D'abord, les consommateurs se plaignent du comportement asymétrique des prix à la pompe principalement lors de chocs du cours du Brent, et la figure 1 montre graphiquement l'existence de chocs importants. Par ailleurs, la volatilité des prix des carburants permet d'obtenir des coefficients plus précis dans notre modèle économétrique par la suite. Les quatre principaux carburants représentés semblent très corrélés au cours du Brent, et relativement proches. La courbe du gazole apparaît au dessus des courbes des autres carburants, montrant un prix au litre HT supérieur au SP95, SP95-E10 et SP98. Le prix hors taxe moyen au litre du gazole est de 1,07 € sur l'année 2022; 0,95 € pour le

1. On procède en convertissant la latitude et longitude de chaque station en radians pour trouver les autres stations à une distance de cinq kilomètres divisés par le rayon moyen de la Terre en kilomètres.

2. Voir [Gautier and Saout \(2015\)](#).

SP95 ; 0,93 € pour le SP95-E10 ; et 1,0 € pour le SP98. Le prix hors taxe moyen au litre du Brent est de 0,60 € pour la période.

**FIGURE 1 – Évolution du prix moyen quotidien HT par carburant en 2022**



Sources : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022), U.S. Energy Information Administration, Boursorama.

La table 1 calcule la corrélation entre le Brent et les différents carburants pour chaque type de station pour l'année 2022. Sans surprise, le E85 et GPLc sont les moins corrélés (respectivement 0,33 et 0,42 sur autoroute). Les autres carburants sont particulièrement corrélés, et confirment l'analyse graphique précédente, ce qui est conforme à la littérature. Le SP95-E10 et SP98 sont les carburants les plus corrélés (plus de 0,9 sur route), suivis par le SP95 et le gazole.

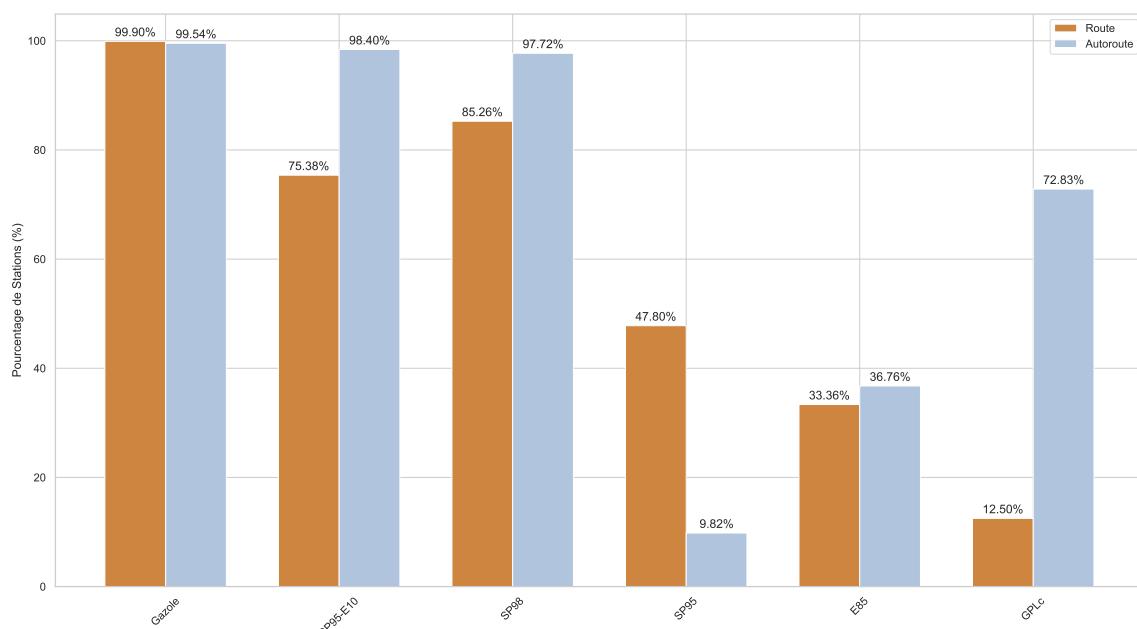
**TABLE 1 – Corrélation entre le cours du Brent et les prix à la pompe HT, par type de carburant et de station-service**

Carburant	Autoroute	Route
Gazole	0.791620	0.778601
SP95-E10	0.902743	0.909465
SP98	0.897102	0.908698
SP95-E5	0.854859	0.904595
E85	0.330473	0.455217
GPLc	0.421298	0.519583

Sources : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022), U.S. Energy Information Administration, Boursorama.

La représentativité et la précision des données pour l'analyse du prix d'un carburant dépendent du nombre de stations-service qui proposent à la vente ce carburant. La figure 2 présente la proportion de stations vendant les différents carburants, selon le type de station. Le gazole apparaît comme le carburant le plus distribué sur le marché français avec plus de 99 % de stations qui en proposent, sur route et autoroute. Il est suivi par le SP95-E10 et SP98 eux aussi largement distribués (plus de 75 % quelque soit le type de station). En revanche, le SP95 n'est vendu que dans 9,82 % de stations d'autoroute. En 15 ans, il a été progressivement remplacé par le SP95-E10, moins coûteux et de qualité équivalente. De ce fait, les données relatives aux prix du SP95 sur autoroute sont particulièrement incomplètes en raison de sa distribution limitée, ce qui pose des problèmes pour la modélisation économétrique.

**FIGURE 2 – Proportion de stations vendant chaque carburant (route et autoroute)**

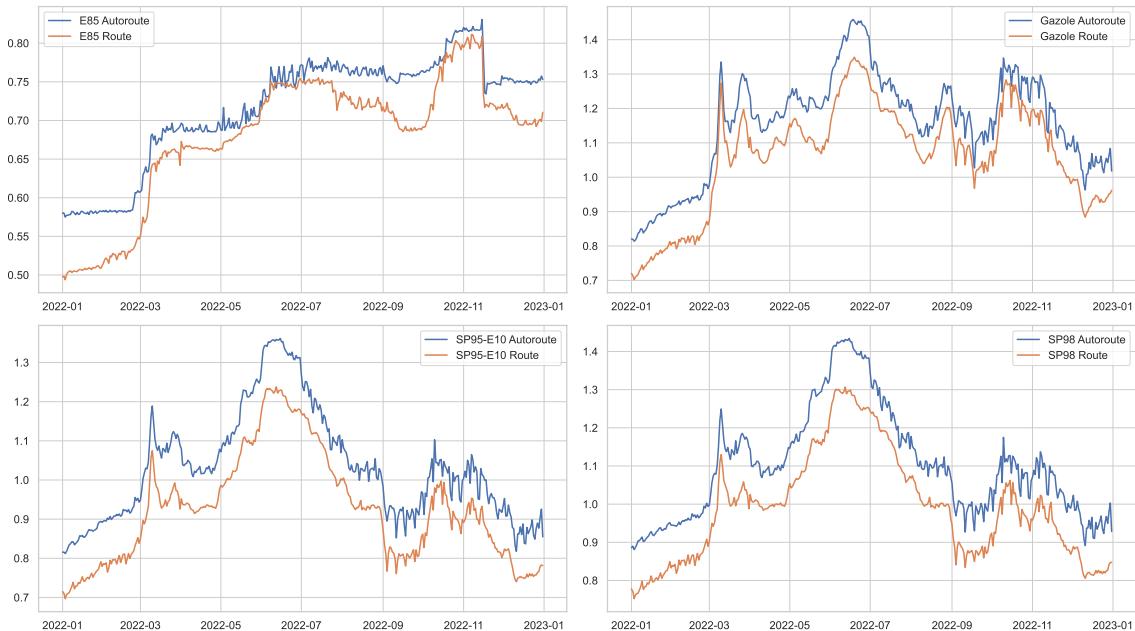


Source : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022).

Les stations-service sur route sont majoritairement des stations de supermarché, contrairement aux stations d'autoroute qui sont gérées sous forme de concessions. Pour pouvoir exploiter une aire d'autoroute, les entreprises doivent s'engager à réaliser des travaux de modernisation, fournir des services spécifiques ou encore assurer la présence d'agents en continu, ce qui justifie des prix plus élevés que sur route. Dans le même temps, ces stations d'autoroute détiennent un pouvoir de marché supérieur aux stations sur route, en raison d'une concurrence géographique moins dense et d'une situation de monopole local (frais de péage pour quitter et revenir sur l'autoroute). L'analyse et la comparaison de ces deux groupes est donc pertinent du point de vue économique.

La figure 3 compare pour quatre carburants les prix moyens quotidiens HT sur l'année 2022, pour les stations de route et les stations d'autoroute (respectivement les courbes orange et bleue). On observe graphiquement que les prix d'autoroute sont plus élevés sur autoroute, pour chaque carburant – E85, gazole, SP95-E10 et SP98. L'écart semble même plutôt stable malgré les variations des prix de chaque type de stations sur la période. Cela concorde effectivement avec la littérature existante.

**FIGURE 3 – Comparaison de l'évolution des prix des carburants : Stations sur route vs autoroute**



Sources : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022), U.S. Energy Information Administration, Boursorama.

La table 2 est un tableau de sortie de test de Student, qui teste l'hypothèse d'égalité des moyennes entre stations sur route et sur autoroute, pour chaque carburant. On rejette l'hypothèse nulle d'égalité pour tous les carburants, conformément aux analyses graphiques précédentes, sauf pour le SP95. Cela est du au nombre restreint de relevés de prix pour les stations d'autoroute (44 sur l'année 2022) qui ne permet pas d'avoir un échantillon suffisant pour étudier les variations du SP95 sur autoroute. Cela témoigne de l'effacement du SP95 au fil des années en faveur du SP95-E10, en particulier sur autoroute.

**TABLE 2 – Test de Student par type de station : Analyse des prix moyens par carburant**

Carburant	Statistique T	P-value	Interprétation
Gazole	8.446482	0.000000	Moyennes significativement différentes.
SP95-E10	10.317616	0.000000	Moyennes significativement différentes.
SP98	10.464462	0.000000	Moyennes significativement différentes.
SP95	1.377391	0.168861	Pas de différence significative.
E85	6.548601	0.000000	Moyennes significativement différentes.
GPLc	26.873514	0.000000	Moyennes significativement différentes.

Source : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022).

La table 3 teste les variances des prix des stations-services sur route face aux stations d'autoroute, et l'hypothèse nulle d'égalité des variances entre ces deux groupes est rejetée pour l'ensemble des carburants. Cela nous éclaire sur le comportement des prix selon le type de station, qui est similaire, ce qui est important pour l'analyse économétrique. Ces groupes sont comparables car de même variance.

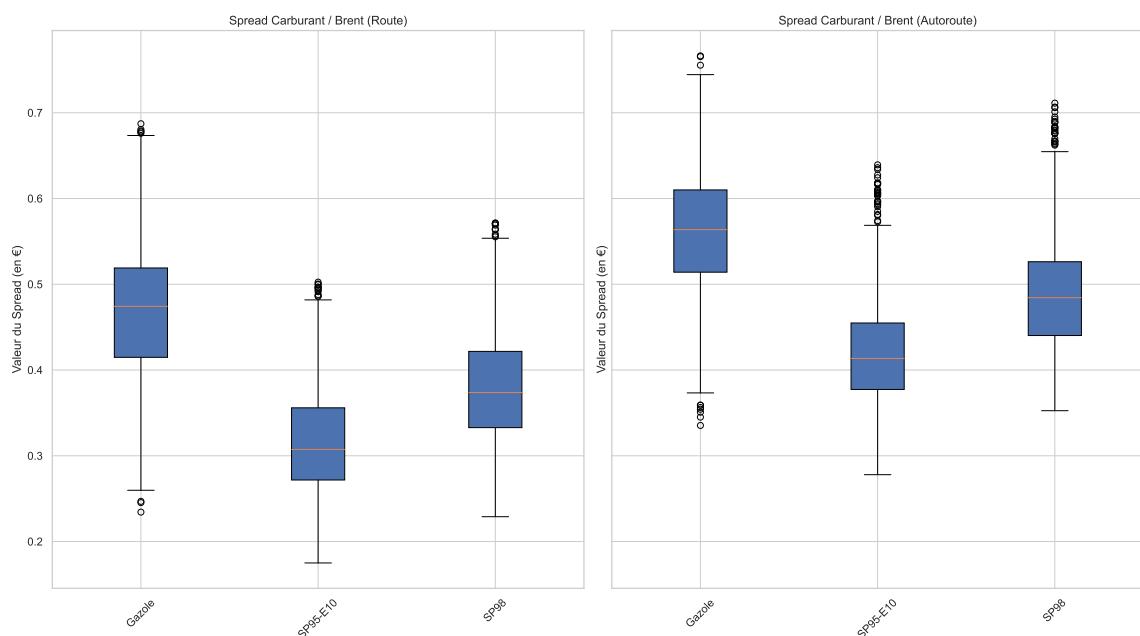
**TABLE 3 – Test de Levene par type de station : Analyse des prix moyens par carburant**

Carburant	Statistique W	P-value	Interprétation
Gazole	0.321226	0.571047	Variances égales entre les groupes.
SP95-E10	0.000727	0.978493	Variances égales entre les groupes.
SP98	0.021499	0.883468	Variances égales entre les groupes.
SP95	0.058831	0.808429	Variances égales entre les groupes.
E85	2.274584	0.131944	Variances égales entre les groupes.
GPLc	3.634755	0.056978	Variances égales entre les groupes.

Source : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022).

La figure 4 est un Box Plot affichant visuellement la distribution des spreads (écart entre le prix HT et le Brent) quotidiens en 2022. Le graphique compare les spreads entre le gazole, le SP95-E10 et le SP98, pour les stations sur route et sur autoroute. L'écart médian entre le prix à la pompe et le prix du Brent est le plus important pour le gazole, suivi du SP98 et du SP95-E10. On observe encore que sur autoroute, l'écart entre les prix à la pompe et le Brent est plus élevé que sur route. Sur autoroute, le spread médian est à 0,55 € pour le gazole ; 0,42 € pour le SP95-E10 ; et 0,48 € pour le SP98. Par ailleurs, on voit graphiquement la présence de valeurs extrêmes en haut (et non en bas) pour le SP95-E10 et SP98 sur route comme sur autoroute. Cela peut indiquer la présence d'asymétrie dans l'ajustement du SP95-E10 et SP98 face aux variations du Brent, et c'est ce que l'on va essayer d'analyser avec notre modèle économétrique par la suite.

**FIGURE 4 – Comparaison de l'évolution des prix des carburants : Stations sur route vs autoroute**



Sources : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022), U.S. Energy Information Administration, Boursorama.

La table 4 présente la durée moyenne des prix des carburants en France en 2022. La durée moyenne est calculée comme l'écart moyen entre deux changements de prix. Pour chaque carburant, la durée moyenne (en jours) est plus importante sur les stations sur route que sur les stations sur autoroute. Ainsi, en 2022, la durée moyenne du prix du gazole sur route était de 2,47 jours contre 1,59 jours sur autoroute. La table met aussi en évidence le cas du SP95 qui ne possède que 43 relevés de prix pour les stations sur autoroute en 2022, ce qui montre que ce carburant n'est quasiment pas vendu sur autoroute. Nous n'inclurons donc pas ce carburant dans notre modèle économétrique qui est marginal dans notre base de données.

Une explication sur cet écart de durée moyenne est que les stations-service sur autoroute sont généralement moins nombreuses et ont moins de concurrence directe, ce qui peut leur permettre de changer les prix plus fréquemment en réponse aux variations du Brent. Une étude plus approfondie sur les changements de prix pourrait déterminer si les ajustements de prix sur autoroute sont significativement plus faible en valeur absolue. [Gautier and Saout \(2015\)](#) trouvent que la durée moyenne des prix du gazole est de 6,6 jours (2007–2009), contre 2,42 jours (en agrégeant les deux types de stations) dans nos données (2022). L'augmentation de la fréquence des ajustements de prix peut être attribuée aux avancées technologiques, en particulier à l'amélioration des communications et à l'accessibilité accrue des informations en ligne, ce qui permet aux gestionnaires de stations-service de réagir plus rapidement aux changements du marché.

**TABLE 4 – Durée moyenne entre deux changements de prix, par carburant et par type de station-service (autoroute ou route)**

Carburant	Type	Nb obs.	Durée moyenne (en jours)	Écart-type	Q25	Q50	Q75
Gazole	Autoroute	436.00	1.59	1.63	1.00	1.00	1.00
	Route	9023.00					
SP95-E5	Autoroute	43.00	3.27	3.02	1.00	2.00	4.00
	Route	4317.00					
SP95-E10	Autoroute	431.00	1.61	1.77	1.00	1.00	1.00
	Route	6808.00					
SP98	Autoroute	428.00	1.63	1.90	1.00	1.00	1.00
	Route	7701.00					

Source : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022).

La table 5 présente le taux de marge moyen par carburant et par type de station. Le taux de marge est obtenu à partir de la moyenne sur l'année des taux de marge quotidiens par carburant et type de station, soit  $1 + \mu = 1 + P_{HT} - P_{Brent}$ . Le terme  $\mu$  correspond au coefficient de marge, soit la différence entre le prix de vente hors taxe  $P_{HT}$  et le coût de production  $P_{Brent}$ . L'expression  $1 + \mu$  peut s'interpréter comme le rapport du prix de vente sur le coût de production. Sans surprise, le taux de marge est supérieur sur autoroute et cela confirme les résultats précédents. Il serait intéressant de comparer les taux de marge obtenus empiriquement par nos relevés de prix, avec le taux de marge obtenu par notre modèle économétrique, et vérifier que les résultats sont cohérents. Il est important de noter que le taux de marge prend en compte le coût de production du pétrole brut, mais pas les étapes intermédiaires de raffinage, de stockage et de transport. Le taux de marge réel est donc inférieur aux valeurs obtenues dans la table 5.

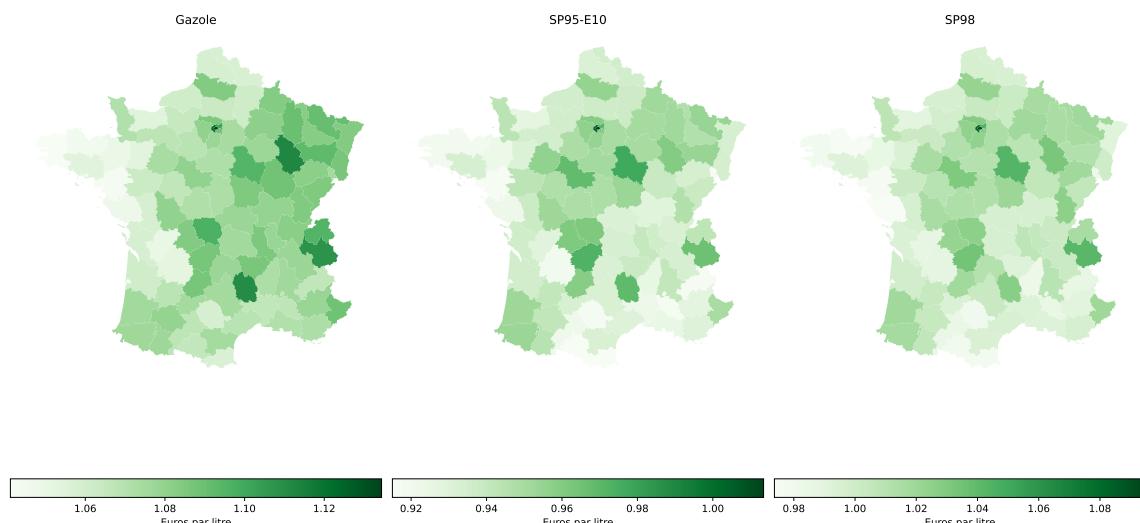
**TABLE 5 – Taux de marge ( $1 + \mu$ ) par carburant et type de station**

Carburant	Route	Autoroute
Gazole	1.761338	1.917962
SP95-E10	1.535205	1.710471
SP98	1.642812	1.823711
SP95	1.580707	1.604844
E85	1.125631	1.187540
GPLc	1.017049	1.165851

Sources : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022), U.S. Energy Information Administration, Boursorama.

La figure 5 est une série de 3 cartes de France représentant chacune les niveaux de prix moyens par département pour les trois principaux carburants en 2022. Pour le gazole, les prix moyens par département vont de 1,06 à 1,12 euros par litre ; de 0,94 à 1,02 euros par litre pour le SP95-E10 ; de 0,98 à 1,08 euros par litre pour le SP98. Les prix à Paris sont toujours les plus élevés, et les prix en Bretagne et Vendée parmi les plus faibles. Néanmoins, les disparités sont plus ou moins fortes selon le type de carburant : le gazole a un prix relativement stable parmi les départements comparativement au SP95-E10. La présence de disparités régionales est intéressante et peut s'expliquer par différents facteurs comme le niveau de concurrence locale, les disparités de pouvoir d'achat ou encore la présence d'autoroutes dans le département. Il serait donc pertinent d'analyser les ajustements des prix selon les départements pour détecter si il existe des écarts notables.

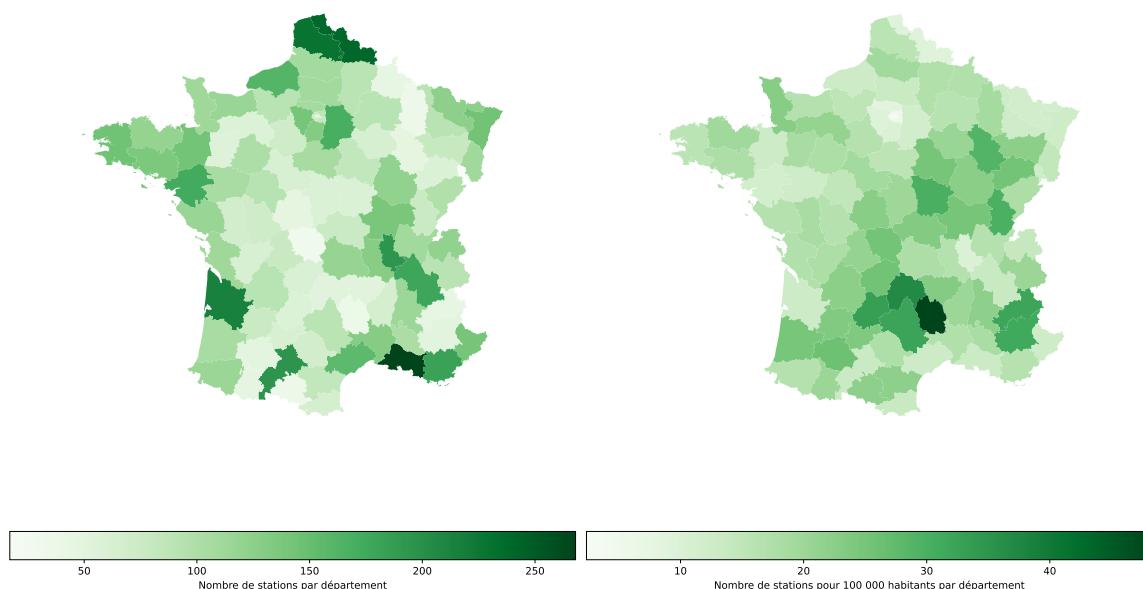
**FIGURE 5 – Disparités régionales des prix des carburants en 2022 : Visualisation par type de carburant (gazole, SP95/E10 et SP98)**



Source : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022).

La figure 6 contient deux cartes : celle de gauche présente le nombre de stations-services par département, que l'on peut interpréter comme le niveau de concurrence géographique par département, tandis que celle de droite rapporte le nombre de stations-service pour 100 000 habitants. La mise en perspective de la carte du nombre de stations-service par département avec la figure 5 permet de constater que les départements avec une forte concentration de stations-service ont des prix relativement plus bas que les départements avec une faible concentration : les départements du Nord et du Pas-de-Calais ainsi que les départements de la côte atlantique ont un fort niveau de concurrence géographique et des prix relativement bas. Cette analyse graphique souligne surtout l'importance de contrôler des variables liées à la concurrence géographique dans notre analyse quantitative, pour ne pas omettre de variables importantes.

**FIGURE 6 – Répartition des station-services françaises : Perspective absolue et relative**



Source : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022).

### 3.3 Méthode utilisée

Bien que de nombreux modèles économétriques différents aient été utilisés pour étudier la présence d'asymétrie dans les variations des prix à la pompe, la littérature économique suggère principalement l'estimation d'un modèle à correction d'erreur. La méthode la plus fréquemment adoptée est celle développée par Engle and Granger (1987), que l'on va utiliser ici. La méthode consiste par estimer une relation de long terme entre les prix des carburants et les prix du pétrole brut. Si les deux séries sont cointégrées, on intègre les résidus de la première relation dans un modèle ECM. On peut dans un troisième temps étudier la présence d'asymétrie. On transforme les prix en log pour réduire l'hétéroscédasticité des résidus, mieux prendre en compte les différences d'échelles, et interpréter les résultats plus facilement.

D'abord, la relation de long terme entre les prix à la pompe hors taxe (HT) et le prix du pétrole brut est donnée par :

$$\ln(PP_t) = \alpha_0 + \alpha_1 \ln(Brent_t) + \alpha_2 Q_2 + \alpha_3 Q_3 + \alpha_4 Q_4 + \epsilon_t \quad (1)$$

avec  $PP_t$  le prix à la pompe HT ;  $Brent_t$  le prix du Brent en euros constants (qui est supposé exogène) ;

$\epsilon_t$  est le terme d'erreur ;  $\alpha_0$  la constante ;  $\alpha_1$  le degré d'ajustement du prix à la pompe en fonction du pétrole brut. L'ajustement à long terme est complet lorsque le coefficient  $\alpha_1$  atteint la valeur 1. Pour  $\alpha_1$  inférieur à 1, cela indique que le marché n'est pas totalement compétitif. Comme avancé par [Rousseas \(1985\)](#),  $\alpha_0$  peut être interprété comme le taux de marge constant. Les variables  $\alpha_2$ ,  $\alpha_3$  et  $\alpha_4$  sont des variables dichotomiques qui prennent la valeur 1 si la station-service fait partie du quartile associé en termes de concurrence géographique. Dit plus simplement, les stations-services ont été réparties en quartiles selon le nombre de stations-service du même type dans un rayon de cinq kilomètres. Cela permet de prendre en compte le niveau de concurrence auquel est confrontée chaque station-service.<sup>3</sup>

Deuxièmement, pour examiner la dynamique à court terme des variations du prix à la pompe du carburant en réponse aux variations du prix du pétrole brut, nous utilisons la méthodologie fournie par [Engle and Granger \(1987\)](#). Dans sa forme standard, le modèle s'écrit donc :

$$\Delta \ln(PP_t) = \sum_{i=0}^2 \beta_{1i} \Delta \ln(Brent_{t-i}) + \beta_2 \Delta \ln(PP_{t-1}) + \theta \hat{\epsilon}_{t-1} + \nu_t \quad (2)$$

avec  $\Delta$  l'opérateur de première différence ; et  $\nu_t$  le terme d'erreur.  $\beta_{1i}$  mesure le taux de transmission de la variation du prix du brut à la période  $(t - i)$  sur le prix à la pompe à la période  $t$  et  $\hat{\epsilon}_{t-1} = (\ln(PP_{t-1}) - \alpha_0 - \alpha_1 \ln(Brent_{t-1}))$  représente le niveau de déséquilibre à la période  $(t - 1)$ , est le résidu de la relation de long terme donnée par l'équation (1) ; et  $\theta$  capture la vitesse d'ajustement de la correction d'erreur lorsque les prix s'éloignent de leur niveau d'équilibre. Pour pouvoir intégrer les résidus de l'équation 2 de la période précédente, il faut avoir testé au préalable la cointégration des cours du pétrole brut et des prix à la pompe.

Pour incorporer la possibilité d'asymétries d'ajustement des prix à la pompe quand ils sont différents de leur niveau de long terme, on segmente le terme à correction d'erreur  $\theta$  de l'équation 2 en deux coefficients,  $\theta^+$  et  $\theta^-$ . On utilise  $\theta^+$  dans le cas où le prix du pétrole brut a augmenté à la période précédente. Inversement, on utilise  $\theta^-$  dans le cas où le prix du pétrole brut a diminué à la période précédente. La dynamique à court terme des variations du prix à la pompe avec ajustement asymétrique est donné par :

$$\Delta \ln(PP_t) = \sum_{i=0}^2 \delta_{1i} \Delta \ln(Brent_{t-i}) + \delta_2 \Delta \ln(PP_{t-1}) + \theta^+ \hat{\epsilon}_{t-1} + \theta^- \hat{\epsilon}_{t-1} + \eta_t \quad (3)$$

avec  $\delta_{1i}$  le taux de transmission de la variation du prix du brut à la période  $(t - i)$  sur le prix à la pompe comme dans le cas symétrique ;  $\theta^+$  capture la vitesse d'ajustement de la correction d'erreur lorsque le prix du Brent a augmenté à la période précédente ;  $\theta^-$  capture la vitesse d'ajustement de la correction d'erreur lorsque le prix du Brent a diminué à la période précédente.

Pour vérifier la présence d'asymétrie, on procède au test de Wald pour vérifier si les coefficients  $\theta^+$  et  $\theta^-$  sont significativement différents.

Dans cette spécification, on étudie l'asymétrie sur la partie dite de long terme du modèle, et non sur les coefficients de court terme. Même si des asymétries peuvent apparaître sur la partie court terme, l'analyse des asymétries de long terme permet d'analyser les causes structurelles et durables des imperfections du marché, ce qui permet d'élaborer des politiques économiques réellement efficaces, c'est pourquoi nous nous concentrerons sur cet aspect dans la suite du mémoire.

---

3. Les stations du premier quartile ont deux concurrents ou moins dans un rayon de cinq kilomètres. Celles du deuxième quartile ont entre deux et quatre concurrents, et entre quatre et neuf concurrents pour le troisième quartile. Les stations du dernier quartile ont plus de neuf concurrents dans un rayon de cinq kilomètres.

## 4 Résultats et discussion

### 4.1 Résultats

Nous suivons donc la méthodologie d'[Engle and Granger \(1987\)](#) pour déterminer la présence d'asymétrie dans les variations des prix des carburants face aux fluctuations du Brent. On commence par modéliser la relation de long terme entre les prix à la pompe et le cours du Brent d'après l'équation 1. On régresse les 3 carburants principaux de notre base de données : gazole, SP95-E10 et SP98. Aussi, on divise les stations en deux groupes : (a) stations sur route, (b) stations sur autoroute. Le nombre d'observations dans la table [6a](#) (route) est supérieur à un million pour chacun des carburants, contre une moyenne de 95 717 observations sur autoroute ([6b](#)).

On constate que la valeur de  $\alpha_1$  est inférieur à 1 pour le gazole, SP95-E10 et SP98 – respectivement 0,86 ; 0,96 ; et 0,92. L'interprétation de  $\alpha_1 < 1$  est que les fluctuations du Brent ne sont pas transmises intégralement vers les différents carburants sur le long terme : le marché n'est pas en situation de concurrence pure et parfaite car une variation d'1 % du brut ne se traduit pas par une variation d'1 % du prix du gazole, SP95-E10 ou SP98. Par ailleurs, le terme constant est supérieur dans la table [6b](#) (autoroute) que celui des stations de route. Cela signifie que le taux de marge constant est plus important pour les stations d'autoroute.

Les prix étant exprimés en logarithme, on peut retrouver le taux de marge "de long terme" des stations, en exponentialisant le terme constant pour chaque carburant, à l'aide de la formule  $1 + \mu = 1 + e^{\alpha_0}$ . Le taux de marge obtenu par le modèle pour le gazole est de 1,63 sur route et 1,72 sur autoroute. En comparaison, le taux de marge calculé lors des statistiques descriptives est de 1,77 et 1,79 (voir [1](#)). Les résultats obtenus par les différentes méthodes sont concordants.

Les tables [6a](#) et [6b](#) modélisent la relation entre les carburants et le Brent, et utilisent pour cela des variables indicatrices relatives à la concurrence géographique. Les stations sont séparées en quatre groupes selon le nombre de stations concurrentes dans un rayon de cinq kilomètres. La catégorie de référence est associée à la présence de deux stations ou moins. Les coefficients  $\alpha_2$ ,  $\alpha_3$  et  $\alpha_4$  mesurent ainsi l'effet relatif d'être dans un de ces quartiles par rapport au premier quartile. Normalement, on s'attend à ce que ces coefficients soient négatifs, indiquant que plus il y a de concurrence, plus les prix tendent à baisser.

Cependant, dans le cas des stations sur route,  $\alpha_2$  n'est pas significatif pour le SP95-E10, ni  $\alpha_3$  pour le SP98. De surcroît, le signe de  $\alpha_4$  – le quartile le plus concurrentiel – est positif pour tous les carburants, ce qui indique que le groupe avec le plus de concurrence a des prix plus élevés plutôt que plus bas.

Cela s'explique par le fait que le rayon tout seul n'est pas suffisant pour expliquer la concurrence des stations sur route. Dans les zones urbaines denses, même si beaucoup de stations sont proches géographiquement, les temps de trajet réels, augmentés par le trafic, peuvent limiter l'effet de la concurrence. De plus, dans les grandes villes comme Paris, où les coûts logistiques et le niveau de vie sont plus élevés, les prix des carburants restent supérieurs malgré une forte présence de concurrents. Ces facteurs viennent brouiller l'effet réel de la hausse de la concurrence sur le prix des carburants pour les stations sur route.

Néanmoins, pour les stations sur autoroute étant soumises à moins d'hétérogénéité spatiale, les valeurs de  $\alpha_2$ ,  $\alpha_3$  et  $\alpha_4$  sont cohérentes avec la théorie. Ainsi, une forte compétition géographique a pour conséquence une baisse de 6 % du prix du gazole sur autoroute, toutes choses égales par ailleurs.

**TABLE 6 – Relation de long-terme entre les prix à la pompe HT et le prix du pétrole brut**

$$\ln(PP_t) = \alpha_0 + \alpha_1 \ln(Brent_t) + \alpha_2 Q_2 + \alpha_3 Q_3 + \alpha_4 Q_4 + \epsilon_t$$

**(a) Stations sur route**

	Gazole	SP95E10	SP98
$\alpha_0$	0.4893*** (0.0004)	0.4004*** (0.0004)	0.4455*** (0.0003)
$\alpha_1$	0.8657*** (0.0007)	0.9625*** (0.0006)	0.9203*** (0.0006)
$\alpha_2$	-0.0027*** (0.0003)	0.0000 (0.0003)	-0.0009*** (0.0003)
$\alpha_3$	-0.0044*** (0.0003)	-0.0006*** (0.0002)	0.0004 (0.0002)
$\alpha_4$	0.0132*** (0.0003)	0.0204*** (0.0002)	0.0241*** (0.0002)
R-squared Adj.	0.5396	0.7008	0.6861
No. observations	1286334	1026934	1021550

**(b) Stations sur autoroute**

	Gazole	SP95E10	SP98
$\alpha_0$	0.5445*** (0.0013)	0.4868*** (0.0013)	0.5267*** (0.0013)
$\alpha_1$	0.7578*** (0.0024)	0.8642*** (0.0024)	0.8161*** (0.0023)
$\alpha_2$	-0.0227*** (0.0008)	-0.0265*** (0.0008)	-0.0269*** (0.0008)
$\alpha_3$	-0.0360*** (0.0010)	-0.0455*** (0.0010)	-0.0423*** (0.0009)
$\alpha_4$	-0.0617*** (0.0012)	-0.0784*** (0.0012)	-0.0751*** (0.0011)
R-squared Adj.	0.5034	0.5899	0.5774
No. observations	98226	95499	93426

Les erreurs types sont indiquées entre parenthèses.

\*\*\*, \*\* et \* indiquent respectivement la significativité aux seuils de 1%, 5% et 10%.

Sources : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022), U.S. Energy Information Administration, Boursorama.

La méthode d'Engle and Granger (1987) est une technique en deux étapes pour tester la cointégration. Premièrement, on estime une relation de long terme. Ensuite, on analyse les résidus de cette régression. Si les résidus de la relation de long terme sont stationnaires, cela signifie que les séries sont cointégrées. Le test augmenté de Dickey-Fuller permet de vérifier la stationnarité des résidus, et de confirmer l'existence d'un équilibre de long terme entre le prix des carburants et le prix du Brent. La table 7 confirme bien la stationnarité des résidus et la cointégration des carburants avec le Brent au seuil de significativité de 1 %. On conclut qu'il existe bien une relation de long terme entre les prix des carburants et le prix du Brent, ce qui nous permet d'estimer un modèle à correction d'erreur.

**TABLE 7 – Test augmenté de Dickey-Fuller sur les résidus de l’équation de long-terme**

Carburant	Type	ADF Statistic	P-valeur	Used Lag	Obs.
Gazole	Route	-264.592493	0.000000	0	1286333
	Autoroute	-58.633813	0.000000	0	98225
SP95E10	Route	-275.567863	0.000000	0	1026933
	Autoroute	-58.246941	0.000000	0	95498
SP98	Route	-280.174678	0.000000	0	1021549
	Autoroute	-56.296158	0.000000	0	93425

Sources : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022), U.S. Energy Information Administration, Boursorama.

Nous pouvons maintenant estimer un modèle à correction d’erreur à partir des résidus de la relation de long terme, d’après l’équation 2. Comme mentionné dans la partie 3.3, le modèle est décomposé en deux parties : une partie de court terme et une de long terme. Le coefficient  $\theta$  capture la vitesse d’ajustement de la correction d’erreur lorsque les prix s’éloignent de leur niveau d’équilibre. On observe dans les tables 8a et 8b que le coefficient  $\theta$  est plus élevé pour le gazole, SP95-E10 et SP98 pour les stations sur route que pour les stations d’autoroute. Dans le cas du SP98, suite à un écart du prix d’équilibre de long terme, environ 8,17 % de cet écart est corrigé chaque période pour les stations sur route. Pour les stations d’autoroute, environ 3,2 % de l’écart entre le prix et le prix de long terme est corrigé chaque période. Un marché parfaitement efficient aurait un  $\theta$  de valeur 1, signifiant que les prix s’ajustent intégralement suite à un écart avec le prix de long terme, ce qui n’est pas le cas du marché des stations-service en France, en 2022.

Les coefficients d’ajustement de court terme sont aussi supérieurs pour les stations sur route que sur autoroute. En effet, sur route, une variation du Brent d’1 % engendre une variation du prix du gazole de 0,169 %, tandis que sur autoroute cela engendre une variation du prix du gazole de 0,06 %.

L’ajustement plus rapide sur le court terme ainsi que sur le long terme pour les stations sur route permet de conclure que le marché des stations-service de route est plus concurrentiel que le marché d’autoroute. C’est aussi ce qu’avaient [Borenstein and Shepard \(2002\)](#) dans leur étude en montrant que l’existence de pouvoir de marché rend les prix des carburants plus rigides. Après avoir analysé les différences de réponse des prix pour les marchés sur route et sur autoroute, on va introduire la possibilité d’ajustements asymétriques dans notre modèle.

**TABLE 8 – Relation de court-terme entre les prix à la pompe HT et le prix du pétrole brut, modèle à correction d’erreur sans asymétrie**

$$\Delta \ln(PP_t) = \sum_{i=0}^2 \beta_{1i} \Delta \ln(Brent_{t-i}) + \beta_2 \Delta \ln(PP_{t-1}) + \theta \hat{\epsilon}_{t-1} + \nu_t$$

**(a) Stations sur route**

	Gazole	SP95E10	SP98
$\beta_{1,0}$	0.1697*** (0.0008)	0.1523*** (0.0008)	0.1748*** (0.0008)
$\beta_{1,1}$	0.2162*** (0.0009)	0.1756*** (0.0009)	0.1729*** (0.0008)
$\beta_{1,2}$	0.1706*** (0.0009)	0.1482*** (0.0009)	0.1372*** (0.0008)
$\beta_2$	0.0074*** (0.0008)	-0.0339*** (0.0010)	-0.0508*** (0.0010)
$\theta$	-0.0752*** (0.0003)	-0.0753*** (0.0004)	-0.0817*** (0.0004)
R-squared Adj.	0.2102	0.1614	0.1726
No. observations	1259283	1006755	998503

**(b) Stations sur autoroute**

	Gazole	SP95E10	SP98
$\beta_{1,0}$	0.0644*** (0.0027)	0.0887*** (0.0026)	0.0846*** (0.0024)
$\beta_{1,1}$	0.1638*** (0.0028)	0.1566*** (0.0026)	0.1431*** (0.0025)
$\beta_{1,2}$	0.1682*** (0.0028)	0.1692*** (0.0027)	0.1570*** (0.0025)
$\beta_2$	0.0112*** (0.0031)	-0.0933*** (0.0032)	-0.0806*** (0.0032)
$\theta$	-0.0486*** (0.0009)	-0.0326*** (0.0008)	-0.0321*** (0.0008)
R-squared Adj.	0.1420	0.1151	0.1144
No. observations	96918	94206	92142

Les erreurs types sont indiquées entre parenthèses.

\*\*\*, \*\* et \* indiquent respectivement la significativité aux seuils de 1%, 5% et 10%.

Sources : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022), U.S. Energy Information Administration, Boursorama.

Le modèle de la table 10 analyse la relation de court terme entre les prix à la pompe hors taxe et les prix du pétrole brut, en intégrant des termes d'asymétrie pour capter les réactions différentes selon que les prix augmentent ou diminuent (équation 3). Le coefficient  $\theta^+$  capture la vitesse d'ajustement du prix actuel par rapport au prix de long terme quand le Brent a augmenté à la période précédente.  $\theta^-$  capture la vitesse d'ajustement du prix actuel par rapport au prix de long terme quand le Brent a diminué à la période précédente. Il est donc possible d'étudier les comportements des prix en situation d'augmentation ou de baisse des prix. On remarque en premier lieu que le coefficient  $\theta^+$  est supérieur au coefficient  $\theta^-$ , et cela pour chaque carburant, pour chaque groupe. Les coefficients sont tous significatifs au seuil de 1 %, et montrent ainsi un ajustement plus fort du prix à la pompe quand le Brent a augmenté à la période précédente, que lorsque le Brent a diminué. Les coefficients  $\theta^+$  et  $\theta^-$  étant issus du coefficient  $\theta$  de la table 8, l'écart de vitesse d'ajustement par rapport au prix de long terme est encore plus rapide pour les stations sur route. Ainsi, pour le gazole sur autoroute, suite à un écart du prix d'équilibre de long terme, environ 8,57 % de cet écart est corrigé à la période t si le Brent a augmenté à la période précédente, contre 6,46 % de l'écart corrigé quand le Brent a diminué à la période précédente.

Pour conclure à la présence d'asymétrie, il faut tester l'hypothèse d'égalité de  $\theta^+$  et  $\theta^-$ , pour chaque carburant et chaque type de station. C'est ce que fait la table 9, et l'hypothèse nulle  $\theta^+ = \theta^-$  est rejetée pour tous les carburants d'après un test de Wald. Il y a donc bien une différence significative dans l'ajustement des prix des carburants en France en 2022, selon si le cours du pétrole brut augmente ou diminue. Les résultats de ce modèle montrent la présence d'asymétrie dans l'ajustement des prix des carburants, affectant tous les types de carburants et de stations-service. Les stations situées sur les routes montrent des coefficients d'ajustement à court et long terme plus élevés que celles sur autoroute, suggérant un niveau de compétition supérieur. Malgré cela, elles présentent également des ajustements asymétriques pour les trois carburants étudiés.

**TABLE 9 – Test de Wald**

Carburant	Type	P-valeur	Interprétation
Gazole	Route	0.000000	$\theta^+ \neq \theta^-$
	Autoroute	0.000000	$\theta^+ \neq \theta^-$
SP95E10	Route	0.000000	$\theta^+ \neq \theta^-$
	Autoroute	0.000000	$\theta^+ \neq \theta^-$
SP98	Route	0.000000	$\theta^+ \neq \theta^-$
	Autoroute	0.000000	$\theta^+ \neq \theta^-$

Sources : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022), U.S. Energy Information Administration, Boursorama.

**TABLE 10 – Relation de court-terme entre les prix à la pompe HT et le prix du pétrole brut, modèle à correction d’erreur avec asymétrie**

$$\Delta \ln(PP_t) = \sum_{i=0}^2 \delta_{1i} \Delta \ln(Brent_{t-i}) + \delta_2 \Delta \ln(PP_{t-1}) + \theta^+ \hat{\epsilon}_{t-1} + \theta^- \hat{\epsilon}_{t-1} + \eta_t$$

**(a) Stations sur route**

	Gazole	SP95E10	SP98
$\delta_{1,0}$	0.1698*** (0.0008)	0.1524*** (0.0008)	0.1749*** (0.0008)
$\delta_{1,1}$	0.2164*** (0.0009)	0.1750*** (0.0009)	0.1726*** (0.0008)
$\delta_{1,2}$	0.1705*** (0.0009)	0.1483*** (0.0009)	0.1372*** (0.0008)
$\delta_2$	0.0057*** (0.0009)	-0.0344*** (0.0010)	-0.0511*** (0.0010)
$\theta^+$	-0.0857*** (0.0004)	-0.0818*** (0.0005)	-0.0854*** (0.0005)
$\theta^-$	-0.0646*** (0.0004)	-0.0695*** (0.0005)	-0.0781*** (0.0005)
R-squared Adj.	0.2111	0.1617	0.1727
No. observations	1259283	1006755	998503

**(b) Stations sur autoroute**

	Gazole	SP95E10	SP98
$\delta_{1,0}$	0.0651*** (0.0027)	0.0887*** (0.0026)	0.0847*** (0.0024)
$\delta_{1,1}$	0.1625*** (0.0028)	0.1560*** (0.0026)	0.1424*** (0.0025)
$\delta_{1,2}$	0.1683*** (0.0028)	0.1692*** (0.0027)	0.1570*** (0.0025)
$\delta_2$	0.0098*** (0.0031)	-0.0930*** (0.0032)	-0.0804*** (0.0032)
$\theta^+$	-0.0619*** (0.0012)	-0.0392*** (0.0013)	-0.0382*** (0.0012)
$\theta^-$	-0.0369*** (0.0012)	-0.0275*** (0.0011)	-0.0274*** (0.0011)
R-squared Adj.	0.1441	0.1156	0.1148
No. observations	96918	94206	92142

Les erreurs types sont indiquées entre parenthèses.

\*\*\*, \*\* et \* indiquent respectivement la significativité aux seuils de 1%, 5% et 10%.

Sources : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022), U.S. Energy Information Administration, Boursorama.

## 4.2 Discussion

Ce mémoire a pour but d'examiner la présence d'asymétrie dans l'ajustement des prix à la pompe français sur la période 2022. Pour ce faire, nous avons étudié le cas des trois carburants les plus consommés en France (gazole, SP95-E10 et SP98), et séparé les stations-service en deux catégories différentes selon leur emplacement : les stations de route, et les stations d'autoroute. Cette méthode se justifie par les différences importantes entre ceux deux groupes. Les stations d'autoroute appartiennent majoritairement à des groupes pétroliers, et bénéficient d'une position monopolistique en raison des barrières à l'entrée (péage) pour accéder à ces voies de circulation. Notre étude a montré que le segment autoroutier a un taux de marge supérieur au segment routier, et des ajustements plus lents. Cela montre le niveau de concurrence moindre sur autoroute ainsi que l'existence de pouvoir de marché plus important pour ce groupe. Malgré la différence de niveau de concurrence, nos résultats indiquent la présence d'asymétrie d'ajustement des prix pour l'ensemble des carburants, et cela indépendamment du groupe étudié.

Nous avons utilisé des données à haute fréquence au niveau microéconomique. La plupart des études à propos de l'asymétrie des prix de carburants utilisent des données agrégées. [Perdiguero-García \(2013\)](#) montre que ces données tendent à minimiser la présence d'asymétrie en raison d'une granularité trop faible. La majorité des études utilisant des données microéconomiques trouvent la présence d'asymétrie : [Remer \(2015\)](#) à propos du marché américain, [Balaguer and Ripollés \(2016\)](#) pour le cas espagnol, ou encore [Balmaceda and Soruco \(2008\)](#) pour le cas chilien. Il n'existe qu'une seule étude sur le cas du marché français utilisant des données à haute fréquence au niveau microéconomique. [Gautier and Saout \(2015\)](#) utilisent la même base de données pour la période 2007–2009 et ne concluent pas à la présence d'asymétrie, et il était donc pertinent de réexaminer avec des données actuelles le cas des stations-service françaises. Nos conclusions diffèrent de celles de [Gautier and Saout \(2015\)](#), qui ne détectent des différences dans les ajustements des prix des carburants pour aucun des carburants. Cela peut s'expliquer par l'écart important entre les deux périodes étudiées – 2007-2009 contre 2022 – ainsi que par les différences dans la spécification des deux modèles. Les auteurs utilisent un modèle empirique qui n'est pas un modèle à correction d'erreur. Le modèle ECM est le plus utilisé dans la littérature et c'est celui que nous avons choisi, ce qui peut expliquer la différence de résultats.

Pour autant, toutes les autres études sur le marché français ont conclu à l'existence d'asymétrie, ce que confirme notre étude. Plus globalement, nos résultats sont en accord avec la littérature économique : [Perdiguero-García \(2013\)](#) examine 403 études différentes sur ce sujet, qui constatent la présence d'asymétrie pour plus de 60% d'entre elles.

## 4.3 Limites

### 4.3.1 Critère de compétition

Notre étude a tenté de mesurer la compétition géographique à laquelle était confrontée chaque station pour étudier son impact sur le prix des carburants. Les stations sur autoroute ont une certaine homogénéité géographique qui permet de mesurer l'impact de la compétition : une station sur autoroute avec plus de neufs concurrents dans un rayon de cinq kilomètres a des prix du gazole plus chers de 6,17 % qu'une station sur autoroute avec deux concurrents ou moins, toutes choses égales par ailleurs. Néanmoins, pour les stations de route, le critère du nombre de stations dans un rayon de cinq kilomètres n'est pas suffisant, comme développé dans la section 4.1. Les zones urbaines denses ont des prix supérieurs en raison d'un pouvoir d'achat plus élevé et de loyers plus chers, alors même que le nombre de stations dans un rayon de cinq kilomètres est supérieur. Pour encore mieux appréhender la concurrence réelle de chaque station, il aurait fallu pouvoir ajouter des variables pour contrôler la

densité démographique, les prix locaux de l’immobilier, ou le revenu moyen par zone. Notons que ces variables présentent des difficultés dans leur construction, en raison notamment des inégalités économiques spatiales au sein de zones urbaines denses par exemple.

De plus, nos données ne contiennent pas d’information sur la marque des station-service. Cette information aurait été pertinente à inclure dans notre étude car toutes les stations n’ont pas la même stratégie commerciale. Par exemple, les stations de supermarché se servent de la vente de carburants comme de prix d’appel pour attirer les consommateurs, tandis que les stations appartenant à des compagnies pétrolières sont dans une logique de maximisation de leur profit grâce à la vente de carburant.

#### 4.3.2 Explications de l’asymétrie

Les données utilisées sont des relevés de prix des stations-services françaises pour l’année 2022. Cela nous permet d’analyser finement les ajustements des différents carburants. Néanmoins, il n’est pas possible à partir de ces données d’établir les étapes dans la chaîne de production et de distribution qui sont source d’asymétrie. Bien que l’hypothèse la plus courante est que les asymétries proviennent du pouvoir de marché de stations-service (côté distribution), il aurait été pertinent de pouvoir analyser l’évolution des coûts de raffinage, stockage et transport pour établir quelles étapes sont les plus génératrices d’asymétries entre l’extraction du pétrole brut et la vente à la pompe. Dans le cas français, [Audenis et al. \(2002\)](#) concluent avec des données mensuelles pour la période 1980–2000 que l’asymétrie provient de chaque étape entre l’extraction du pétrole jusqu’à la vente au détail.

## 5 Conclusion

Utilisant des données à l’échelle microéconomique issues de 4 537 449 relevés de prix pour l’année 2022 en France, nous avons montré l’existence d’asymétries significatives dans les réponses des prix des carburants face aux variations du cours du Brent. Les prix s’ajustent plus rapidement lors d’une hausse du prix du Brent que lors d’une baisse. Ces résultats sont valables pour les trois carburants étudiés (gazole, SP95-E10 et SP98) et les deux types de stations-services (route, autoroute). Le secteur des stations-service sur route apparaît plus concurrentiel que le marché des stations-service sur autoroute. Nos conclusions sont en accord avec les travaux sur ce sujet utilisant des données à l’échelle de stations-service, renforçant la littérature sur les asymétries de prix à haute fréquence, avec une contribution supplémentaire. Nos résultats montrent que le marché de la vente de carburants en France n’est pas en situation de concurrence pure et parfaite, car les fluctuations du Brent ne sont pas transmises intégralement vers les différents carburants sur le long terme, d’après notre régression de long terme. Plusieurs pistes de politiques publiques pourraient diminuer le pouvoir de marché des stations et les corriger les asymétries observées. Premièrement, le développement d’outils numériques pour améliorer la transparence des prix et faire baisser le coût de recherche des consommateurs permettrait de diminuer les asymétries des prix liées à la demande, et stimulerait la concurrence. L’augmentation des contrôles et de la surveillance vis-à-vis des pratiques de collusion et d’ententes entre stations-service tendrait à réduire les pratiques monopolistiques et l’asymétrie de réponse des prix des carburants. Enfin, l’introduction de plafonnement des prix à la pompe en cas de chocs internationaux contribuerait à limiter les effets d’asymétrie constatés.

## Bibliographie

- Audenis, C., Biscourp, P., and Riedinger, N. (2002). Le prix des carburants est plus sensible à une hausse qu'à une baisse du brut. *Economie et Statistique*, 359(1) :149–165.
- Bachmeier, L. J. and Griffin, J. M. (2003). New Evidence on Asymmetric Gasoline Price Responses. *The Review of Economics and Statistics*, 85(3) :772–776.
- Bacon, R. W. (1991). Rockets and feathers : the asymmetric speed of adjustment of UK retail gasoline prices to cost changes. *Energy Economics*, 13(3) :211–218.
- Balaguer, J. and Ripollés, J. (2016). Asymmetric fuel price responses under heterogeneity. *Energy Economics*, 54 :281–290.
- Balmaceda, F. and Soruco, P. (2008). Asymmetric Dynamic Pricing in a Local Gasoline Retail Market\*. *The Journal of Industrial Economics*, 56(3) :629–653.
- Bettendorf, L., van der Geest, S. A., and Varkevisser, M. (2003). Price asymmetry in the Dutch retail gasoline market. *Energy Economics*, 25(6) :669–689.
- Borenstein, S., Cameron, A. C., and Gilbert, R. (1997). Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes ? *The Quarterly Journal of Economics*, 112(1) :305–339.
- Borenstein, S. and Shepard, A. (2002). Sticky Prices, Inventories, and Market Power in Wholesale Gasoline Markets. *The RAND Journal of Economics*, 33(1) :116–139.
- Boroumand, R. H., Goutte, S., Porcher, S., and Porcher, T. (2016). Asymmetric evidence of gasoline price responses in France : A Markov-switching approach. *Economic Modelling*, 52 :467–476.
- Boroumand, R. H., Porcher, T., and Urom, C. (2021). Negative oil price shocks transmission : The comparative effects of the GFC, shale oil boom, and Covid-19 downturn on French gasoline prices. *Research in International Business and Finance*, 58 :101455.
- Deltas, G. (2008). Retail Gasoline Price Dynamics and Local Market Power\*. *The Journal of Industrial Economics*, 56(3) :613–628.
- Ederington, L. H., Fernando, C. S., Hoelscher, S. A., Lee, T. K., and Linn, S. C. (2019). A review of the evidence on the relation between crude oil prices and petroleum product prices. *Journal of Commodity Markets*, 13 :1–15.
- Engle, R. F. and Granger, C. W. J. (1987). Co-Integration and Error Correction : Representation, Estimation, and Testing. *Econometrica*, 55(2) :251–276.
- Frey, G. and Manera, M. (2007). Econometric Models of Asymmetric Price Transmission. *Journal of Economic Surveys*, 21(2) :349–415.
- Gautier, E. and Saout, R. L. (2015). The Dynamics of Gasoline Prices : Evidence from Daily French Micro Data. *Journal of Money, Credit and Banking*, 47(6) :1063–1089.
- Grasso, M. and Manera, M. (2007). Asymmetric error correction models for the oil–gasoline price relationship. *Energy Policy*, 35(1) :156–177.
- Hosken, D. S., McMillan, R. S., and Taylor, C. T. (2008). Retail gasoline pricing : What do we know ? *International Journal of Industrial Organization*, 26(6) :1425–1436.
- Johnson, R. N. (2002). Search Costs, Lags and Prices at the Pump. *Review of Industrial Organization*, 20(1) :33–50.
- Karrenbrock, J. D. (1991). The Behavior of Retail Gasoline Prices : Symmetric or Not? *Review*, 73(4).
- Kilian, L. (2010). Explaining Fluctuations in Gasoline Prices : A Joint Model of the Global Crude Oil Market and the U.S. Retail Gasoline Market. *The Energy Journal*, 31(2) :87–112.
- Lamotte, O., Porcher, T., Schalck, C., and Silvestre, S. (2013). Asymmetric gasoline price responses in France. *Applied Economics Letters*, 20(5) :457–461.
- Leparmetier, A. (2023). Joe Biden condamne les profits excessifs, sans pouvoir y faire grand-chose. *Le Monde.fr*.

- Lewis, M. S. (2011). Asymmetric Price Adjustment and Consumer Search : An Examination of the Retail Gasoline Market. *Journal of Economics & Management Strategy*, 20(2) :409–449.
- Liu, M.-H., Margaritis, D., and Tourani-Rad, A. (2010). Is there an asymmetry in the response of diesel and petrol prices to crude oil price changes ? Evidence from New Zealand. *Energy Economics*, 32(4) :926–932.
- Manning, D. N. (1991). Petrol prices, oil price rises and oil price falls : some evidence for the UK since 1972. *Applied Economics*, 23(9) :1535–1541.
- Misak, E. (2023). Les prix des produits pétroliers en 2022 : des niveaux inédits liés à la crise géopolitique.
- Perdiguero-García, J. (2013). Symmetric or asymmetric oil prices ? A meta-analysis approach. *Energy Policy*, 57 :389–397.
- Reagan, P. B. and Weitzman, M. L. (1982). Asymmetries in price and quantity adjustments by the competitive firm. *Journal of Economic Theory*, 27(2) :410–420.
- Remer, M. (2015). An empirical investigation of the determinants of asymmetric pricing. *International Journal of Industrial Organization*, 42 :46–56.
- Rousseas, S. (1985). A Markup Theory of Bank Loan Rates. *Journal of Post Keynesian Economics*.
- Shin, D. (1994). Do product prices respond symmetrically to changes in crude prices ? *OPEC Review*, 18 :137–157.
- Tappata, M. (2009). Rockets and feathers : Understanding asymmetric pricing. *The RAND Journal of Economics*, 40(4) :673–687.
- Yang, H. and Ye, L. (2008). Search with learning : understanding asymmetric price adjustments. *The RAND Journal of Economics*, 39(2) :547–564.

## Annexes

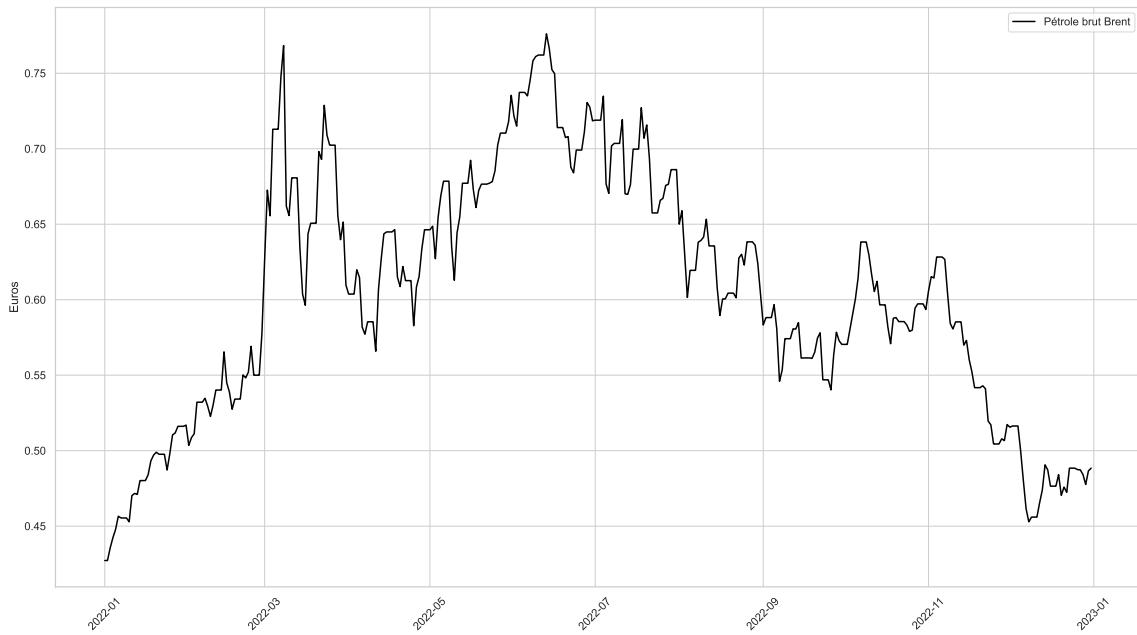
### 5.1

Lien vers une carte interactive que j'ai codée sur Python avec la librairie Folium, représentant l'ensemble des stations-services présentes dans la base de données.

[Cliquez ici pour accéder au lien Github :](#)

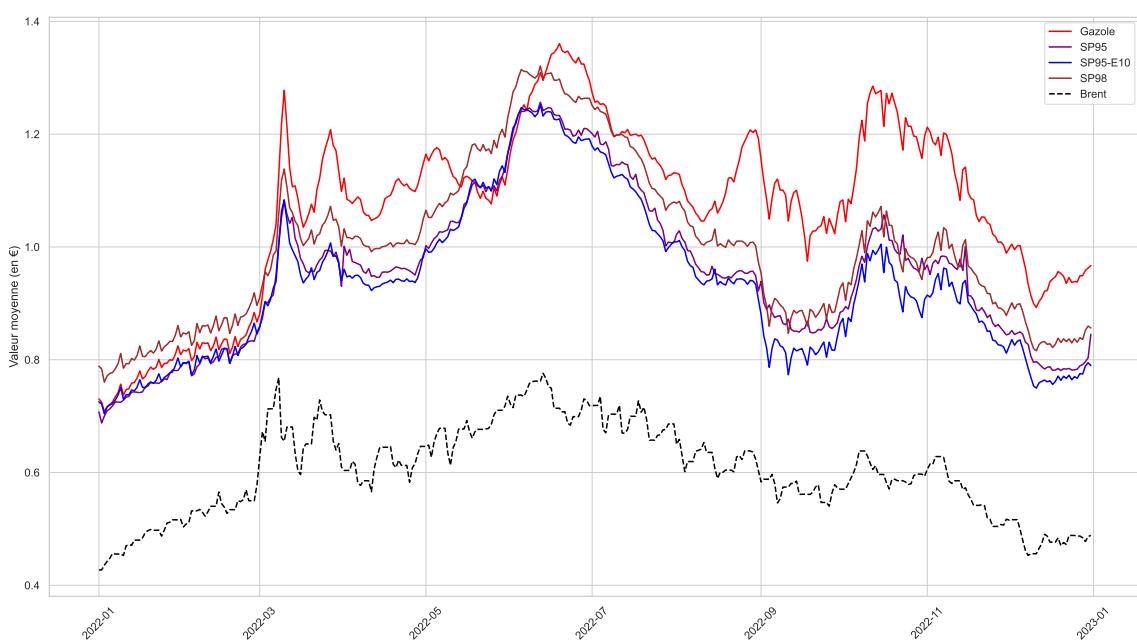
## 5.2

**FIGURE 7 – Évolution du prix moyen quotidien du Brent en euros en 2022**



Sources : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022), U.S. Energy Information Administration, Boursorama.

**FIGURE 8 – Évolution du prix moyen quotidien HT par carburant en 2022 (E85 et GPLc non inclus)**



Sources : Données quotidiennes provenant du site gouvernemental [prix-carburants.gouv.fr](http://prix-carburants.gouv.fr) (1<sup>er</sup> janvier - 31 décembre 2022), U.S. Energy Information Administration, Boursorama.

## Code Python

La structure du code suit cette présentation :

1. **Préparation de la base de données** : reformatage des coordonnées GPS et les dates, récupération des valeurs quotidiennes du Brent en euros pour chaque relevé de prix, passage aux prix HT, mise en place de la variable de compétition géographique.
2. **Statistiques descriptives** : graphiques, tests statistiques, tableaux.
3. **Modèle économétrique** : relation de long terme, modèle ECM sans asymétrie, modèle ECM avec asymétrie, test ADF et test de Wald.

# v1\_coordonnées

May 7, 2024

```
[1]: import os
os.chdir("C:/Users/zetru/OneDrive/Mémoire python/")
import numpy as np
```

```
[2]: import pandas as pd

chemin_du_fichier = 'points-extremes-des-departements-metropolitains-de-france.
↪csv'

df_coordonnées = pd.read_csv(chemin_du_fichier)

print(df_coordonnées.head())
```

	Departement	Latitude la plus au nord	Latitude la plus au sud \
0	01	46.517199	45.611235
1	02	50.069272	48.837795
2	03	46.803872	45.930728
3	04	44.659501	43.668282
4	05	45.126844	44.186479

	Longitude la plus à l'est	Longitude la plus à l'ouest
0	6.169736	4.729097
1	4.254064	2.962451
2	4.005570	2.280403
3	6.966820	5.498010
4	7.077105	5.418533

```
[3]: df_coordonnées
```

```
[3]: Departement  Latitude la plus au nord  Latitude la plus au sud \
0          01            46.517199            45.611235
1          02            50.069272            48.837795
2          03            46.803872            45.930728
3          04            44.659501            43.668282
4          05            45.126844            44.186479
..          ...
91         93            49.012398            48.807438
92         94            48.861405            48.688327
```

```

93      95          49.232197          48.908679
94      2A          42.381405          41.362165
95      2B          43.011724          41.832144

```

	Longitude la plus à l'est	Longitude la plus à l'ouest
0	6.169736	4.729097
1	4.254064	2.962451
2	4.005570	2.280403
3	6.966820	5.498010
4	7.077105	5.418533
..	...	...
91	2.602600	2.288254
92	2.613652	2.310222
93	2.590528	1.608799
94	9.407322	8.540103
95	9.559226	8.573409

[96 rows x 5 columns]

```
[4]: df = pd.read_hdf('df.h5', 'df')
df
```

```

[4]:      id type      latitude      longitude code_postal \
0    1000001   R      4620100      519800     01000
1    1000001   R      4620100      519800     01000
2    1000001   R      4620100      519800     01000
3    1000001   R      4620100      519800     01000
4    1000001   R      4620100      519800     01000
..    ... ...
5048601 95870010   R  4893688.8607759  220698.30018219    95870
5048602 95870010   R  4893688.8607759  220698.30018219    95870
5048603 95870010   R  4893688.8607759  220698.30018219    95870
5048604 95870010   R  4893688.8607759  220698.30018219    95870
5048605 96600001   R      4875637      241048     94600

```

```

            ville nom_carburant id_carburant \
0  SAINT-DENIS-Lès-BOURG      Gazole      1
1  SAINT-DENIS-Lès-BOURG      Gazole      1
2  SAINT-DENIS-Lès-BOURG      Gazole      1
3  SAINT-DENIS-Lès-BOURG      Gazole      1
4  SAINT-DENIS-Lès-BOURG      Gazole      1
..    ...
5048601           Bezons      SP98       6
5048602           Bezons      SP98       6
5048603           Bezons      SP98       6
5048604           Bezons      SP98       6
5048605  Choisy-le-Roi      None    None

```

```

                maj    valeur
0      2022-01-03T08:44:18  1.572
1      2022-01-05T10:08:11  1.560
2      2022-01-10T10:09:08  1.603
3      2022-01-17T12:34:42  1.616
4      2022-01-17T12:34:44  1.616
...
5048601   ...   ...
5048602   2022-12-27T20:42:31  1.790
5048603   2022-12-28T18:50:37  1.790
5048604   2022-12-29T19:28:13  1.796
5048604   2022-12-30T18:38:23  1.804
5048605           None     NaN

```

[5048606 rows x 10 columns]

```

[5]: df['cp'] = df['code_postal'].astype(str).str[:2]

[6]: df_merge = pd.merge(df, df_coordonnées, left_on='cp', right_on='Departement')

[7]: df_merge.drop(df_merge.columns[-5], axis=1, inplace=True)

[8]: def transform_longitude(value):
    value_no_dot = value.replace('.', '')
    if value_no_dot.startswith('-'):
        return value_no_dot[:5]
    else:
        return value_no_dot[:4]

def transform_latitude(value):
    value_no_dot = value.replace('.', '')
    return value_no_dot[:5]

df_merge['latitude'] = df_merge['latitude'].apply(transform_latitude)
df_merge['longitude'] = df_merge['longitude'].apply(transform_longitude)

df_merge

```

```

[8]:      id type latitude longitude code_postal          ville \
0      1000001   R    46201      5198      01000  SAINT-DENIS-LÈS-BOURG
1      1000001   R    46201      5198      01000  SAINT-DENIS-LÈS-BOURG
2      1000001   R    46201      5198      01000  SAINT-DENIS-LÈS-BOURG
3      1000001   R    46201      5198      01000  SAINT-DENIS-LÈS-BOURG
4      1000001   R    46201      5198      01000  SAINT-DENIS-LÈS-BOURG
...
...   ...   ...
5032426  95870010   R    48936      2206      95870          Bezons
5032427  95870010   R    48936      2206      95870          Bezons

```

5032428	95870010	R	48936	2206	95870	Bezons
5032429	95870010	R	48936	2206	95870	Bezons
5032430	95870010	R	48936	2206	95870	Bezons

	nom_carburant	id_carburant		maj	valeur	cp	\
0	Gazole	1	2022-01-03T08:44:18	1.572	01		
1	Gazole	1	2022-01-05T10:08:11	1.560	01		
2	Gazole	1	2022-01-10T10:09:08	1.603	01		
3	Gazole	1	2022-01-17T12:34:42	1.616	01		
4	Gazole	1	2022-01-17T12:34:44	1.616	01		
...	...	...	...	...	...	..	
5032426	SP98	6	2022-12-23T14:47:11	1.768	95		
5032427	SP98	6	2022-12-27T20:42:31	1.790	95		
5032428	SP98	6	2022-12-28T18:50:37	1.790	95		
5032429	SP98	6	2022-12-29T19:28:13	1.796	95		
5032430	SP98	6	2022-12-30T18:38:23	1.804	95		

	Latitude la plus au nord	Latitude la plus au sud	\
0	46.517199	45.611235	
1	46.517199	45.611235	
2	46.517199	45.611235	
3	46.517199	45.611235	
4	46.517199	45.611235	
...	...	...	
5032426	49.232197	48.908679	
5032427	49.232197	48.908679	
5032428	49.232197	48.908679	
5032429	49.232197	48.908679	
5032430	49.232197	48.908679	

	Longitude la plus à l'est	Longitude la plus à l'ouest
0	6.169736	4.729097
1	6.169736	4.729097
2	6.169736	4.729097
3	6.169736	4.729097
4	6.169736	4.729097
...	...	...
5032426	2.590528	1.608799
5032427	2.590528	1.608799
5032428	2.590528	1.608799
5032429	2.590528	1.608799
5032430	2.590528	1.608799

[5032431 rows x 15 columns]

```
[9]: import pandas as pd
```

```

def adjust_value(val_str, lower_bound, upper_bound):
    is_negative = val_str.startswith('-')
    prefix = '-' if is_negative else ''
    val = val_str[1:] if is_negative else val_str
    modifications = ['', '0.', '0.0']
    for mod in modifications:
        test_val = prefix + mod + val if mod else prefix + val[:1] + '.' + val[1:]
    try:
        test_float = float(test_val)
        if lower_bound <= test_float <= upper_bound:
            return test_val
    except ValueError:
        continue
    return val_str

def adjust_group_longitude(group):
    representative_row = group.iloc[0]
    val_str = str(representative_row['longitude'])
    lower_bound = representative_row['Longitude la plus à l'ouest']
    upper_bound = representative_row['Longitude la plus à l'est']
    adjusted_val = adjust_value(val_str, lower_bound, upper_bound)
    return adjusted_val

adjusted_longitudes = df_merge.groupby('id').apply(adjust_group_longitude)

df_merge['adjusted_longitude'] = df_merge['id'].map(adjusted_longitudes)

```

```
[10]: df_merge['adjusted_latitude']=df_merge['latitude'].str[:2] +'.' +df_merge['latitude'].str[2:]
```

```
[11]: df_merge.info()
```

#	Column	Dtype
0	id	object
1	type	object
2	latitude	object
3	longitude	object
4	code_postal	object
5	ville	object
6	nom_carburant	object
7	id_carburant	object
8	maj	object

```

9    valeur           float64
10   cp              object
11   Latitude la plus au nord   float64
12   Latitude la plus au sud    float64
13   Longitude la plus à l'est   float64
14   Longitude la plus à l'ouest  float64
15   adjusted_longitude        object
16   adjusted_latitude         object
dtypes: float64(5), object(12)
memory usage: 652.7+ MB

```

```
df_final=df_merge.drop(columns=['Latitude la plus au nord','Latitude la plus au sud','Longitude la plus à l'est','Longitude la plus à l'ouest']) df_final
```

[12]: df\_final=df\_merge

[13]: cols = list(df\_final.columns)  
last\_col,ante\_col = cols[-1],cols[-2]  
cols.remove(last\_col)  
cols.remove(ante\_col)  
cols.insert(2, last\_col)  
cols.insert(3,ante\_col)  
df\_final = df\_final[cols]  
df\_final

		id	type	adjusted_latitude	adjusted_longitude	latitude	longitude	code_postal	ville	nom_carburant
0		1000001	R	46.201		5.198	46201		SAINT-DENIS-LÈS-BOURG	Gazole
1		1000001	R	46.201		5.198	46201		SAINT-DENIS-LÈS-BOURG	Gazole
2		1000001	R	46.201		5.198	46201		SAINT-DENIS-LÈS-BOURG	Gazole
3		1000001	R	46.201		5.198	46201		SAINT-DENIS-LÈS-BOURG	Gazole
4		1000001	R	46.201		5.198	46201		SAINT-DENIS-LÈS-BOURG	Gazole
...		...	...	...		...	...	...	...	...
5032426	95870010		R	48.936		2.206	48936			
5032427	95870010		R	48.936		2.206	48936			
5032428	95870010		R	48.936		2.206	48936			
5032429	95870010		R	48.936		2.206	48936			
5032430	95870010		R	48.936		2.206	48936			
...	...	...	...	...		...	...	...	...	...
5032426		2206		95870		Bezons				SP98
5032427		2206		95870		Bezons				SP98
5032428		2206		95870		Bezons				SP98

5032429	2206	95870	Bezons	SP98
5032430	2206	95870	Bezons	SP98
0	id_carburant	maj	valeur	cp \
1	1	2022-01-03T08:44:18	1.572	01
2	1	2022-01-05T10:08:11	1.560	01
3	1	2022-01-10T10:09:08	1.603	01
4	1	2022-01-17T12:34:42	1.616	01
...	...	...	..	
5032426	6	2022-12-23T14:47:11	1.768	95
5032427	6	2022-12-27T20:42:31	1.790	95
5032428	6	2022-12-28T18:50:37	1.790	95
5032429	6	2022-12-29T19:28:13	1.796	95
5032430	6	2022-12-30T18:38:23	1.804	95
0	Latitude la plus au nord	Latitude la plus au sud	\	
1	46.517199	45.611235		
2	46.517199	45.611235		
3	46.517199	45.611235		
4	46.517199	45.611235		
...	...	...		
5032426	49.232197	48.908679		
5032427	49.232197	48.908679		
5032428	49.232197	48.908679		
5032429	49.232197	48.908679		
5032430	49.232197	48.908679		
0	Longitude la plus à l'est	Longitude la plus à l'ouest		
1	6.169736	4.729097		
2	6.169736	4.729097		
3	6.169736	4.729097		
4	6.169736	4.729097		
...	...	...		
5032426	2.590528	1.608799		
5032427	2.590528	1.608799		
5032428	2.590528	1.608799		
5032429	2.590528	1.608799		
5032430	2.590528	1.608799		

[5032431 rows x 17 columns]

[14]: df\_final=df\_final.drop(columns=['latitude','longitude'])

[15]: df\_final=df\_final.dropna()

```
[16]: df_final['adjusted_latitude'] = pd.to_numeric(df_final['adjusted_latitude'])
df_final['adjusted_longitude'] = pd.to_numeric(df_final['adjusted_longitude'])

# Affichage des informations du DataFrame pour vérifier les types de données
df_final
```

```
[16]:      id type  adjusted_latitude  adjusted_longitude code_postal \
0       1000001    R             46.201            5.198     01000
1       1000001    R             46.201            5.198     01000
2       1000001    R             46.201            5.198     01000
3       1000001    R             46.201            5.198     01000
4       1000001    R             46.201            5.198     01000
...
...   ...
5032426  95870010    R             48.936            2.206     95870
5032427  95870010    R             48.936            2.206     95870
5032428  95870010    R             48.936            2.206     95870
5032429  95870010    R             48.936            2.206     95870
5032430  95870010    R             48.936            2.206     95870

      ville nom_carburant id_carburant \
0  SAINT-DENIS-Lès-BOURG      Gazole        1
1  SAINT-DENIS-Lès-BOURG      Gazole        1
2  SAINT-DENIS-Lès-BOURG      Gazole        1
3  SAINT-DENIS-Lès-BOURG      Gazole        1
4  SAINT-DENIS-Lès-BOURG      Gazole        1
...
...   ...
5032426          Bezons      SP98         6
5032427          Bezons      SP98         6
5032428          Bezons      SP98         6
5032429          Bezons      SP98         6
5032430          Bezons      SP98         6

      maj  valeur  cp  Latitude la plus au nord \
0  2022-01-03T08:44:18    1.572  01           46.517199
1  2022-01-05T10:08:11    1.560  01           46.517199
2  2022-01-10T10:09:08    1.603  01           46.517199
3  2022-01-17T12:34:42    1.616  01           46.517199
4  2022-01-17T12:34:44    1.616  01           46.517199
...
...   ...
5032426  2022-12-23T14:47:11    1.768  95           49.232197
5032427  2022-12-27T20:42:31    1.790  95           49.232197
5032428  2022-12-28T18:50:37    1.790  95           49.232197
5032429  2022-12-29T19:28:13    1.796  95           49.232197
5032430  2022-12-30T18:38:23    1.804  95           49.232197

      Latitude la plus au sud  Longitude la plus à l'est \
0                  45.611235            6.169736
```

```

1           45.611235      6.169736
2           45.611235      6.169736
3           45.611235      6.169736
4           45.611235      6.169736
...
          ...
5032426     48.908679      2.590528
5032427     48.908679      2.590528
5032428     48.908679      2.590528
5032429     48.908679      2.590528
5032430     48.908679      2.590528

    Longitude la plus à l'ouest
0           4.729097
1           4.729097
2           4.729097
3           4.729097
4           4.729097
...
          ...
5032426     1.608799
5032427     1.608799
5032428     1.608799
5032429     1.608799
5032430     1.608799

```

[5028439 rows x 15 columns]

```
[17]: df_final=df_final.rename(columns={'adjusted_latitude':
                                     'latitude','adjusted_longitude':'longitude'})
```

```
[18]: # j'enlève les quelques stations qui n'ont pas des coordonnées cohérentes : 5
       stations soit 837 occurrences.
```

```
[19]: df_final[(df_final['latitude']<df_final['Latitude la plus au\u2022
          \u2022sud'])|(df_final['latitude']>df_final['Latitude la plus au\u2022
          \u2022nord'])|(df_final['longitude']<df_final['Longitude la plus à\u2022
          \u2022l'ouest'])|(df_final['longitude']>df_final['Longitude la plus à\u2022
          \u2022l'est'])]['id'].nunique()
```

```
[19]: 5
```

```
[20]: df_final=df_final[~((df_final['latitude']<df_final['Latitude la plus au\u2022
          \u2022sud'])|(df_final['latitude']>df_final['Latitude la plus au\u2022
          \u2022nord'])|(df_final['longitude']<df_final['Longitude la plus à\u2022
          \u2022l'ouest'])|(df_final['longitude']>df_final['Longitude la plus à l'est']))]
```

```
[21]: 5028439-len(df_final)
```

[21]: 837

```
[22]: df_final=df_final.drop(columns=['Latitude la plus au nord','Latitude la plus au sud','Longitude la plus à l'est','Longitude la plus à l'ouest'])
```

```
df_final
```

```
[22]:      id type  latitude  longitude code_postal \
0     1000001    R     46.201     5.198     01000
1     1000001    R     46.201     5.198     01000
2     1000001    R     46.201     5.198     01000
3     1000001    R     46.201     5.198     01000
4     1000001    R     46.201     5.198     01000
...   ... ...
5032426 95870010    R     48.936     2.206     95870
5032427 95870010    R     48.936     2.206     95870
5032428 95870010    R     48.936     2.206     95870
5032429 95870010    R     48.936     2.206     95870
5032430 95870010    R     48.936     2.206     95870

      ville nom_carburant id_carburant \
0  SAINT-DENIS-Lès-BOURG      Gazole        1
1  SAINT-DENIS-Lès-BOURG      Gazole        1
2  SAINT-DENIS-Lès-BOURG      Gazole        1
3  SAINT-DENIS-Lès-BOURG      Gazole        1
4  SAINT-DENIS-Lès-BOURG      Gazole        1
...   ...
5032426           Bezons      SP98         6
5032427           Bezons      SP98         6
5032428           Bezons      SP98         6
5032429           Bezons      SP98         6
5032430           Bezons      SP98         6

      maj  valeur  cp
0  2022-01-03T08:44:18  1.572  01
1  2022-01-05T10:08:11  1.560  01
2  2022-01-10T10:09:08  1.603  01
3  2022-01-17T12:34:42  1.616  01
4  2022-01-17T12:34:44  1.616  01
...   ...
5032426 2022-12-23T14:47:11  1.768  95
5032427 2022-12-27T20:42:31  1.790  95
5032428 2022-12-28T18:50:37  1.790  95
5032429 2022-12-29T19:28:13  1.796  95
5032430 2022-12-30T18:38:23  1.804  95
```

[5027602 rows x 11 columns]

```
[23]: df_v1=df_final
```

```
[24]: df_v1.to_hdf('df_v1.h5',key='df_v1',mode='w')
```

## v2\_date

May 7, 2024

```
[1]: import os  
os.chdir("C:/Users/zetru/OneDrive/Mémoire python/")  
import numpy as np  
import pandas as pd
```

```
[2]: df=pd.read_hdf('df_v1.h5','df_v1')  
df
```

```
[2]:      id type  latitude  longitude code_postal \
0     1000001     R    46.201     5.198     01000
1     1000001     R    46.201     5.198     01000
2     1000001     R    46.201     5.198     01000
3     1000001     R    46.201     5.198     01000
4     1000001     R    46.201     5.198     01000
...   ...   ...   ...   ...
5032426 95870010     R    48.936     2.206     95870
5032427 95870010     R    48.936     2.206     95870
5032428 95870010     R    48.936     2.206     95870
5032429 95870010     R    48.936     2.206     95870
5032430 95870010     R    48.936     2.206     95870

           ville nom_carburant id_carburant \
0  SAINT-DENIS-Lès-BOURG      Gazole        1
1  SAINT-DENIS-Lès-BOURG      Gazole        1
2  SAINT-DENIS-Lès-BOURG      Gazole        1
3  SAINT-DENIS-Lès-BOURG      Gazole        1
4  SAINT-DENIS-Lès-BOURG      Gazole        1
...   ...   ...   ...
5032426          Bezons      SP98         6
5032427          Bezons      SP98         6
5032428          Bezons      SP98         6
5032429          Bezons      SP98         6
5032430          Bezons      SP98         6

      maj valeur cp
0  2022-01-03T08:44:18  1.572  01
1  2022-01-05T10:08:11  1.560  01
2  2022-01-10T10:09:08  1.603  01
```

```

3      2022-01-17T12:34:42  1.616  01
4      2022-01-17T12:34:44  1.616  01
...
5032426 2022-12-23T14:47:11  1.768  95
5032427 2022-12-27T20:42:31  1.790  95
5032428 2022-12-28T18:50:37  1.790  95
5032429 2022-12-29T19:28:13  1.796  95
5032430 2022-12-30T18:38:23  1.804  95

```

[5027602 rows x 11 columns]

[3]: df['maj']=pd.to\_datetime(df['maj']).dt.date  
df['maj']=pd.to\_datetime(df['maj'])

[4]: df['maj'].info()

```

<class 'pandas.core.series.Series'>
Index: 5027602 entries, 0 to 5032430
Series name: maj
Non-Null Count    Dtype
-----
5027602 non-null   datetime64[ns]
dtypes: datetime64[ns](1)
memory usage: 76.7 MB

```

[5]: df

	id	type	latitude	longitude	code_postal	\		
0	1000001	R	46.201	5.198	01000			
1	1000001	R	46.201	5.198	01000			
2	1000001	R	46.201	5.198	01000			
3	1000001	R	46.201	5.198	01000			
4	1000001	R	46.201	5.198	01000			
...	...	...	...	...	...			
5032426	95870010	R	48.936	2.206	95870			
5032427	95870010	R	48.936	2.206	95870			
5032428	95870010	R	48.936	2.206	95870			
5032429	95870010	R	48.936	2.206	95870			
5032430	95870010	R	48.936	2.206	95870			
			ville	nom_carburant	id_carburant	maj	valeur	\
0	SAINT-DENIS-Lès-BOURG		Gazole		1	2022-01-03	1.572	
1	SAINT-DENIS-Lès-BOURG		Gazole		1	2022-01-05	1.560	
2	SAINT-DENIS-Lès-BOURG		Gazole		1	2022-01-10	1.603	
3	SAINT-DENIS-Lès-BOURG		Gazole		1	2022-01-17	1.616	
4	SAINT-DENIS-Lès-BOURG		Gazole		1	2022-01-17	1.616	
...	...	...	...	...	...	...	...	
5032426	Bezons		SP98		6	2022-12-23	1.768	

```
5032427           Bezons      SP98      6 2022-12-27  1.790
5032428           Bezons      SP98      6 2022-12-28  1.790
5032429           Bezons      SP98      6 2022-12-29  1.796
5032430           Bezons      SP98      6 2022-12-30  1.804

          cp
0       01
1       01
2       01
3       01
4       01
...
5032426   95
5032427   95
5032428   95
5032429   95
5032430   95
```

[5027602 rows x 11 columns]

[6]: df\_v2=df

[7]: df\_v2.to\_hdf('df\_v2.h5',key='df\_v2',mode='w')

## v4\_prixHT\_&\_brent\_par\_litre

May 7, 2024

```
[1]: import os
os.chdir("C:/Users/zetru/OneDrive/Mémoire python/")
import numpy as np
import pandas as pd
df=pd.read_hdf('df_v3.h5','df_v3')

[2]: import matplotlib.pyplot as plt
%matplotlib inline
plt.rcParams['figure.figsize'] = 16, 9
plt.rcParams['axes.grid'] = True

[3]: df['litre_brut']=df['Prix_Brent_euros']/158.987294928

[4]: df.drop(columns=['Data 1: Europe Brent Spot Price FOB (Dollars per Barrel)', 'Prix_Brent_euros', 'Date'], inplace=True)

[5]: df
```

	cp	id	nom_carburant	maj	type	latitude	longitude	\
0	01	1000001	Gazole	2022-01-03	R	46.201	5.198	
1	01	1000001	Gazole	2022-01-05	R	46.201	5.198	
2	01	1000001	Gazole	2022-01-10	R	46.201	5.198	
3	01	1000001	Gazole	2022-01-17	R	46.201	5.198	
4	01	1000001	Gazole	2022-01-21	R	46.201	5.198	
...	...	...	...	...	...	...	...	
4537459	95	95870010	SP98	2022-12-23	R	48.936	2.206	
4537460	95	95870010	SP98	2022-12-27	R	48.936	2.206	
4537461	95	95870010	SP98	2022-12-28	R	48.936	2.206	
4537462	95	95870010	SP98	2022-12-29	R	48.936	2.206	
4537463	95	95870010	SP98	2022-12-30	R	48.936	2.206	
	code_postal		ville	id_carburant	valeur	litre_brut		
0	01000	SAINT-DENIS-LÈS-BOURG		1	1.572	0.435626		
1	01000	SAINT-DENIS-LÈS-BOURG		1	1.560	0.447949		
2	01000	SAINT-DENIS-LÈS-BOURG		1	1.603	0.452874		
3	01000	SAINT-DENIS-LÈS-BOURG		1	1.616	0.484043		
4	01000	SAINT-DENIS-LÈS-BOURG		1	1.652	0.497616		
...	...	...	...	...	...	...		

```

4537459      95870      Bezons      6  1.768  0.488413
4537460      95870      Bezons      6  1.790  0.487324
4537461      95870      Bezons      6  1.790  0.484175
4537462      95870      Bezons      6  1.796  0.477600
4537463      95870      Bezons      6  1.804  0.486593

```

[4537464 rows x 12 columns]

```

[6]: ticpe={
    'Gazole':0.6075,
    'SP95':0.6902,
    'SP98':0.6902,
    'E10':0.6702,
    'E85':0.1183,
    'GPLc':0.2071
}

ticpe_idf={
    'Gazole':0.6264,
    'SP95':0.7004,
    'SP98':0.7004,
    'E10':0.6804,
    'E85':0.1183,
    'GPLc':0.2071
}
tva=0.2
idf=['75','77','78','91','92','93','94','95']

```

```

[7]: df['valeur'] = np.where(
    df['cp'].isin(idf),
    (df['valeur'] / (1 + tva)) - df['nom_carburant'].map(ticpe_idf),
    (df['valeur'] / (1 + tva)) - df['nom_carburant'].map(ticpe)
)

```

```

[8]: remise15=pd.to_datetime(['2022-04-01','2022-08-31'])
remise25=pd.to_datetime(['2022-09-01','2022-11-15'])
remise25

```

```
[8]: DatetimeIndex(['2022-09-01', '2022-11-15'], dtype='datetime64[ns]', freq=None)
```

```

[9]: df.loc[(df['maj']>=remise15[0])&(df['maj']<=remise15[1]), 'valeur'] += 0.15
df.loc[(df['maj']>=remise25[0])&(df['maj']<=remise25[1]), 'valeur'] += 0.25
df.loc[(df['maj']>remise25[1]), 'valeur'] += 0.0833

```

```

[10]: valeurs_aberrantes=df.
    ~index[~((df['nom_carburant']=='E85')|(df['nom_carburant']=='GPLc'))&(df['valeur']<df['litre']
df=df.drop(valeurs_aberrantes).reset_index(drop=True)

```

```
[11]: df_stats=pd.DataFrame(df.groupby(['nom_carburant','maj'])['valeur'].mean())
df_stats.reset_index(inplace=True)
```

```
[13]: df
```

```
[13]:      cp      id nom_carburant      maj type  latitude longitude \
0      01  1000001      Gazole 2022-01-03     R   46.201    5.198
1      01  1000001      Gazole 2022-01-05     R   46.201    5.198
2      01  1000001      Gazole 2022-01-10     R   46.201    5.198
3      01  1000001      Gazole 2022-01-17     R   46.201    5.198
4      01  1000001      Gazole 2022-01-21     R   46.201    5.198
...   ...
4537444  95  95870010      SP98 2022-12-23     R   48.936    2.206
4537445  95  95870010      SP98 2022-12-27     R   48.936    2.206
4537446  95  95870010      SP98 2022-12-28     R   48.936    2.206
4537447  95  95870010      SP98 2022-12-29     R   48.936    2.206
4537448  95  95870010      SP98 2022-12-30     R   48.936    2.206

      code_postal           ville id_carburant      valeur litre_brut
0          01000  SAINT-DENIS-LÈS-BOURG      1  0.702500  0.435626
1          01000  SAINT-DENIS-LÈS-BOURG      1  0.692500  0.447949
2          01000  SAINT-DENIS-LÈS-BOURG      1  0.728333  0.452874
3          01000  SAINT-DENIS-LÈS-BOURG      1  0.739167  0.484043
4          01000  SAINT-DENIS-LÈS-BOURG      1  0.769167  0.497616
...   ...
4537444      95870           Bezons      6  0.856233  0.488413
4537445      95870           Bezons      6  0.874567  0.487324
4537446      95870           Bezons      6  0.874567  0.484175
4537447      95870           Bezons      6  0.879567  0.477600
4537448      95870           Bezons      6  0.886233  0.486593
```

[4537449 rows x 12 columns]

```
[14]: col_temp=df.pop(df.columns[-1])
df.insert(10,col_temp.name,col_temp)
```

```
[15]: col_temp=df.pop(df.columns[-1])
df.insert(4,col_temp.name,col_temp)
```

```
[16]: df.drop('code_postal',axis=1,inplace=True)
```

```
[17]: df_temp=df.pop('litre_brut')
df.insert(5,'litre_brut',df_temp)
```

```
[18]: df_temp=df.pop('id_carburant')
df.insert(3,'id_carburant',df_temp)
```

```
[19]: df
```

```
[19]:      cp      id nom_carburant id_carburant      maj     valeur \
0      01  1000001      Gazole           1 2022-01-03  0.702500
1      01  1000001      Gazole           1 2022-01-05  0.692500
2      01  1000001      Gazole           1 2022-01-10  0.728333
3      01  1000001      Gazole           1 2022-01-17  0.739167
4      01  1000001      Gazole           1 2022-01-21  0.769167
...
4537444  95  95870010          ...        SP98           ...   ...
4537445  95  95870010          ...        SP98           ...   ...
4537446  95  95870010          ...        SP98           ...   ...
4537447  95  95870010          ...        SP98           ...   ...
4537448  95  95870010          ...        SP98           ...   ...

      litre_brut type  latitude longitude            ville
0      0.435626    R    46.201    5.198  SAINT-DENIS-LÈS-BOURG
1      0.447949    R    46.201    5.198  SAINT-DENIS-LÈS-BOURG
2      0.452874    R    46.201    5.198  SAINT-DENIS-LÈS-BOURG
3      0.484043    R    46.201    5.198  SAINT-DENIS-LÈS-BOURG
4      0.497616    R    46.201    5.198  SAINT-DENIS-LÈS-BOURG
...
4537444  0.488413    R    48.936    2.206           ...
4537445  0.487324    R    48.936    2.206           ...
4537446  0.484175    R    48.936    2.206           ...
4537447  0.477600    R    48.936    2.206           ...
4537448  0.486593    R    48.936    2.206           ...

[4537449 rows x 11 columns]
```

```
[20]: df['id'].value_counts()
```

```
[20]: id
33167001    2036
65429001    1842
53200005    1825
44460001    1825
56700004    1825
...
53100006      3
38510004      2
43200008      2
37400008      2
78180012      1
Name: count, Length: 9470, dtype: int64
```

```
[21]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4537449 entries, 0 to 4537448
Data columns (total 11 columns):
 #   Column           Dtype  
 --- 
 0   cp               object  
 1   id               object  
 2   nom_carburant   object  
 3   id_carburant    object  
 4   maj              datetime64[ns]
 5   valeur           float64 
 6   litre_brut      float64 
 7   type             object  
 8   latitude          float64 
 9   longitude         float64 
 10  ville             object  
dtypes: datetime64[ns](1), float64(4), object(6)
memory usage: 380.8+ MB
```

```
[23]: df.to_hdf('df_v4.h5',key='df_v4',mode='w')
```

# v5\_competition\_géographique

May 7, 2024

```
[1]: import os
os.chdir("C:/Users/zetru/OneDrive/Mémoire python/")
import numpy as np
import pandas as pd
import geopandas as gpd
import matplotlib.pyplot as plt
df=pd.read_hdf('df_v4.h5','df_v4')
df
```

```
[1]:      cp      id nom_carburant id_carburant      maj     valeur \
0      01  1000001        Gazole           1 2022-01-03  0.702500
1      01  1000001        Gazole           1 2022-01-05  0.692500
2      01  1000001        Gazole           1 2022-01-10  0.728333
3      01  1000001        Gazole           1 2022-01-17  0.739167
4      01  1000001        Gazole           1 2022-01-21  0.769167
...   ...
4537444  95  95870010       SP98           ...       ...  0.856233
4537445  95  95870010       SP98           ...       ...  0.874567
4537446  95  95870010       SP98           ...       ...  0.874567
4537447  95  95870010       SP98           ...       ...  0.879567
4537448  95  95870010       SP98           ...       ...  0.886233

      litre_brut type  latitude  longitude          ville
0      0.435626    R    46.201     5.198  SAINT-DENIS-LÈS-BOURG
1      0.447949    R    46.201     5.198  SAINT-DENIS-LÈS-BOURG
2      0.452874    R    46.201     5.198  SAINT-DENIS-LÈS-BOURG
3      0.484043    R    46.201     5.198  SAINT-DENIS-LÈS-BOURG
4      0.497616    R    46.201     5.198  SAINT-DENIS-LÈS-BOURG
...   ...
4537444  0.488413    R    48.936     2.206          ...
4537445  0.487324    R    48.936     2.206          ...
4537446  0.484175    R    48.936     2.206          ...
4537447  0.477600    R    48.936     2.206          ...
4537448  0.486593    R    48.936     2.206          ...

[4537449 rows x 11 columns]
```

```
[2]: from scipy.spatial import cKDTree

def degrees_to_radians(df):
    df.loc[:, 'latitude_rad'] = np.radians(df['latitude'])
    df.loc[:, 'longitude_rad'] = np.radians(df['longitude'])
    return df

def calculate_competition(unique_df, all_df, radius_km, station_type=['A', ↴'R']):
    radius_in_radians = radius_km / 6371.0088 # Rayon moyen de la terre en km

    # Filtrer les données uniques pour ne garder que les types spécifiés
    filtered_unique_df = unique_df[unique_df['type'].isin(station_type)]

    # Conversion des degrés en radians
    filtered_unique_df = degrees_to_radians(filtered_unique_df)
    coordinates = filtered_unique_df[['latitude_rad', 'longitude_rad']].values

    # Création de l'arbre KD à partir des coordonnées filtrées
    tree = cKDTree(coordinates)

    # Préparation des données uniques pour l'ensemble des stations
    unique_df = degrees_to_radians(unique_df)
    all_coordinates = unique_df[['latitude_rad', 'longitude_rad']].values

    # Recherche des voisins dans le rayon pour chaque point dans l'ensemble ↴filtré
    counts = tree.query_ball_point(all_coordinates, r=radius_in_radians)

    # Calcul du nombre de stations du même type dans le rayon
    unique_df[f'stations_rayon_{radius_km}km'] = [len(count) - 1 for count in counts]

    all_df = all_df.merge(unique_df[['id', f'stations_rayon_{radius_km}km']], ↴on='id', how='left')

    return all_df

unique_stations = df[['id', 'latitude', 'longitude', 'type']].drop_duplicates(subset='id')

df_stations = calculate_competition(unique_stations, df, 5)
```

```
[3]: df_stations
```

```
[3]: cp      id nom_carburant id_carburant      maj      valeur \
0       01    1000001        Gazole           1 2022-01-03  0.702500
```

```

1      01  1000001      Gazole      1 2022-01-05  0.692500
2      01  1000001      Gazole      1 2022-01-10  0.728333
3      01  1000001      Gazole      1 2022-01-17  0.739167
4      01  1000001      Gazole      1 2022-01-21  0.769167
...
4537444  95  95870010      SP98      6 2022-12-23  0.856233
4537445  95  95870010      SP98      6 2022-12-27  0.874567
4537446  95  95870010      SP98      6 2022-12-28  0.874567
4537447  95  95870010      SP98      6 2022-12-29  0.879567
4537448  95  95870010      SP98      6 2022-12-30  0.886233

      litre_brut type  latitude longitude           ville \
0        0.435626   R    46.201     5.198  SAINT-DENIS-Lès-BOURG
1        0.447949   R    46.201     5.198  SAINT-DENIS-Lès-BOURG
2        0.452874   R    46.201     5.198  SAINT-DENIS-Lès-BOURG
3        0.484043   R    46.201     5.198  SAINT-DENIS-Lès-BOURG
4        0.497616   R    46.201     5.198  SAINT-DENIS-Lès-BOURG
...
4537444  0.488413   R    48.936     2.206          ...
4537445  0.487324   R    48.936     2.206          Bezons
4537446  0.484175   R    48.936     2.206          Bezons
4537447  0.477600   R    48.936     2.206          Bezons
4537448  0.486593   R    48.936     2.206          Bezons

      stations_rayon_5km
0            5
1            5
2            5
3            5
4            5
...
4537444        ...
4537445        15
4537446        15
4537447        15
4537448        15

[4537449 rows x 12 columns]

```

[6]: df\_stations['stations\_rayon\_5km'].quantile(0.75)

[6]: 9.0

[27]: labels = [1, 2, 3, 4]  
df\_stations['competition\_geo'] = pd.qcut(df\_stations['stations\_rayon\_5km'],  
q=4, labels=labels)

```
[28]: df_stations
```

```
[28]:      cp      id nom_carburant id_carburant      maj     valeur \
0      01  1000001      Gazole           1 2022-01-03  0.702500
1      01  1000001      Gazole           1 2022-01-05  0.692500
2      01  1000001      Gazole           1 2022-01-10  0.728333
3      01  1000001      Gazole           1 2022-01-17  0.739167
4      01  1000001      Gazole           1 2022-01-21  0.769167
...
4537444  95  95870010          ...        SP98           6 2022-12-23  0.856233
4537445  95  95870010          ...        SP98           6 2022-12-27  0.874567
4537446  95  95870010          ...        SP98           6 2022-12-28  0.874567
4537447  95  95870010          ...        SP98           6 2022-12-29  0.879567
4537448  95  95870010          ...        SP98           6 2022-12-30  0.886233

      litre_brut type  latitude longitude           ville \
0      0.435626    R    46.201   5.198  SAINT-DENIS-LÈS-BOURG
1      0.447949    R    46.201   5.198  SAINT-DENIS-LÈS-BOURG
2      0.452874    R    46.201   5.198  SAINT-DENIS-LÈS-BOURG
3      0.484043    R    46.201   5.198  SAINT-DENIS-LÈS-BOURG
4      0.497616    R    46.201   5.198  SAINT-DENIS-LÈS-BOURG
...
4537444  0.488413    R    48.936   2.206           ...
4537445  0.487324    R    48.936   2.206           ...
4537446  0.484175    R    48.936   2.206           ...
4537447  0.477600    R    48.936   2.206           ...
4537448  0.486593    R    48.936   2.206           ...

      stations_rayon_5km competition_geo
0                  5            3
1                  5            3
2                  5            3
3                  5            3
4                  5            3
...
4537444          ...
4537445          15           4
4537446          15           4
4537447          15           4
4537448          15           4

[4537449 rows x 13 columns]
```

```
[29]: df_stations['stations_rayon_5km'].idxmax()
```

```
[29]: 376568
```

```
[30]: df_a=df_stations[df_stations['type']=='A']
df_r=df_stations[df_stations['type']=='R']
```

```
[31]: df_a['stations_rayon_5km'].describe()
```

```
[31]: count    380270.000000
mean      3.636705
std       3.967912
min      0.000000
25%     1.000000
50%     2.000000
75%     4.000000
max     23.000000
Name: stations_rayon_5km, dtype: float64
```

```
[32]: df_r['stations_rayon_5km'].describe()
```

```
[32]: count    4.157179e+06
mean      6.420914e+00
std       6.229963e+00
min      0.000000e+00
25%     2.000000e+00
50%     4.000000e+00
75%     9.000000e+00
max     3.600000e+01
Name: stations_rayon_5km, dtype: float64
```

```
[33]: df_stations.sample(10)
```

```
[33]:   cp      id nom_carburant id_carburant      maj    valeur \
3125462  69  69330006        Gazole           1 2022-12-08  0.999967
541296   14  14100007         E10            5 2022-04-29  0.903967
3661539   78  78230002         E85            3 2022-07-28  0.780867
1784459   40  40700004        SP95           2 2022-06-03  1.267300
2177595   51  51000001         E85            3 2022-01-22  0.464200
2384551   56  56860004         E85            3 2022-04-12  0.672533
4015320   85  85150001        Gazole          1 2022-01-04  0.638333
774670    21  21490003         E85            3 2022-05-07  0.697533
2031041   45  45520003         E10            5 2022-03-10  1.224800
2531846   59  59310001         E85            3 2022-10-13  0.880867

      litre_brut type  latitude longitude           ville \
3125462     0.452857   R    45.756     5.0670      PUSIGNAN
541296      0.646327   R    49.133     0.2245      LISIEUX
3661539      0.676483   R    48.881     2.0970      Le Pecq
1784459      0.737303   R    43.657    -0.5800     Hagetmau
2177595      0.497616   R    48.942     4.3890 CHÂLONS-EN-CHAMPAGNE
```

2384551	0.606473	R	47.650	-2.7220	SENE
4015320	0.442323	R	46.607	-1.6640	LES ACHARDS
774670	0.678488	A	47.422	5.1700	BROGNON
2031041	0.655740	A	47.980	1.8560	GIDY
2531846	0.612185	R	50.470	3.2460	Orchies
		stations_rayon_5km	competition_geo		
3125462		2	1		
541296		3	2		
3661539		11	4		
1784459		2	1		
2177595		3	2		
2384551		6	3		
4015320		1	1		
774670		1	1		
2031041		3	2		
2531846		3	2		

```
[34]: df_stations['competition_geo']=df_stations['competition_geo'].astype(object)
```

```
[35]: df_stations.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4537449 entries, 0 to 4537448
Data columns (total 13 columns):
 #   Column           Dtype    
--- 
 0   cp               object    
 1   id               object    
 2   nom_carburant   object    
 3   id_carburant   object    
 4   maj              datetime64[ns]
 5   valeur          float64  
 6   litre_brut      float64  
 7   type             object    
 8   latitude         float64  
 9   longitude        float64  
 10  ville            object    
 11  stations_rayon_5km  int64    
 12  competition_geo  object    
dtypes: datetime64[ns](1), float64(4), int64(1), object(7)
memory usage: 450.0+ MB
```

```
[36]: df_stations.to_hdf('df_v5.h5',key='df_v5',mode='w')
```

```
C:\Users\zetru\AppData\Local\Temp\ipykernel_31972\3041895132.py:1:
PerformanceWarning:
your performance may suffer as PyTables will pickle object types that it cannot
```

```
map directly to c-types [inferred_type->mixed-integer,key->block3_values]
[items->Index(['cp', 'id', 'nom_carburant', 'id_carburant', 'type', 'ville',
   'competition_geo'],
   dtype='object')]

df_stations.to_hdf('df_v5.h5',key='df_v5',mode='w')
```

## graphiques\_on\_v5

May 7, 2024

```
[1]: import os
os.chdir("C:/Users/zetru/OneDrive/Mémoire python/")
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
plt.rcParams['figure.figsize'] = 16, 9
plt.rcParams['axes.grid'] = True
import seaborn as sns
sns.set_theme(style="whitegrid")
```

```
[2]: import os
os.chdir("C:/Users/zetru/OneDrive/Mémoire python/")
import numpy as np
import pandas as pd
df=pd.read_hdf('df_v5.h5','df_v5')
df
```

```
[2]:      cp      id nom_carburant id_carburant      maj      valeur \
0       01  1000001        Gazole           1 2022-01-03  0.702500
1       01  1000001        Gazole           1 2022-01-05  0.692500
2       01  1000001        Gazole           1 2022-01-10  0.728333
3       01  1000001        Gazole           1 2022-01-17  0.739167
4       01  1000001        Gazole           1 2022-01-21  0.769167
...
4537444  95  95870010        SP98           6 2022-12-23  0.856233
4537445  95  95870010        SP98           6 2022-12-27  0.874567
4537446  95  95870010        SP98           6 2022-12-28  0.874567
4537447  95  95870010        SP98           6 2022-12-29  0.879567
4537448  95  95870010        SP98           6 2022-12-30  0.886233

      litre_brut type  latitude longitude          ville \
0        0.435626    R     46.201     5.198  SAINT-DENIS-Lès-BOURG
1        0.447949    R     46.201     5.198  SAINT-DENIS-Lès-BOURG
2        0.452874    R     46.201     5.198  SAINT-DENIS-Lès-BOURG
3        0.484043    R     46.201     5.198  SAINT-DENIS-Lès-BOURG
4        0.497616    R     46.201     5.198  SAINT-DENIS-Lès-BOURG
...
```

```
4537444    0.488413    R    48.936    2.206      Bezons
4537445    0.487324    R    48.936    2.206      Bezons
4537446    0.484175    R    48.936    2.206      Bezons
4537447    0.477600    R    48.936    2.206      Bezons
4537448    0.486593    R    48.936    2.206      Bezons
```

```
stations_rayon_5km competition_geo
0                  5            3
1                  5            3
2                  5            3
3                  5            3
4                  5            3
...
4537444        15           4
4537445        15           4
4537446        15           4
4537447        15           4
4537448        15           4
```

[4537449 rows x 13 columns]

```
[3]: df['nom_carburant'] = df['nom_carburant'].replace({'E10': 'SP95-E10'})
```

```
[4]: import matplotlib.pyplot as plt
%matplotlib inline
plt.rcParams['figure.figsize'] = 16, 9
plt.rcParams['axes.grid'] = True
```

```
[10]: df.groupby(['maj'])['litre_brut'].mean().mean()
```

```
[10]: 0.6035232204907219
```

```
[5]: df_stats=pd.DataFrame(df.groupby(['nom_carburant','maj'])['valeur'].mean())
df_stats
```

```
[5]:          valeur
nom_carburant maj
E85           2022-01-01  0.507524
              2022-01-02  0.509262
              2022-01-03  0.500251
              2022-01-04  0.507993
              2022-01-05  0.510227
...
SP98          2022-12-27  0.839136
              2022-12-28  0.835993
              2022-12-29  0.852364
              2022-12-30  0.859754
```

2022-12-31 0.856226

[2190 rows x 1 columns]

```
[6]: df_stats.reset_index(inplace=True)
```

```
[7]: df_brent=pd.DataFrame(df.groupby(['maj'])['litre_brut'].mean())
df_brent.reset_index(inplace=True)
```

```
[8]: couleurs_carburants = {
    'Gazole': 'red',
    'SP95': 'purple',
    'SP98': 'brown',
    'E85': 'orange',
    'GPLc': 'green',
    'SP95-E10': 'blue'
}

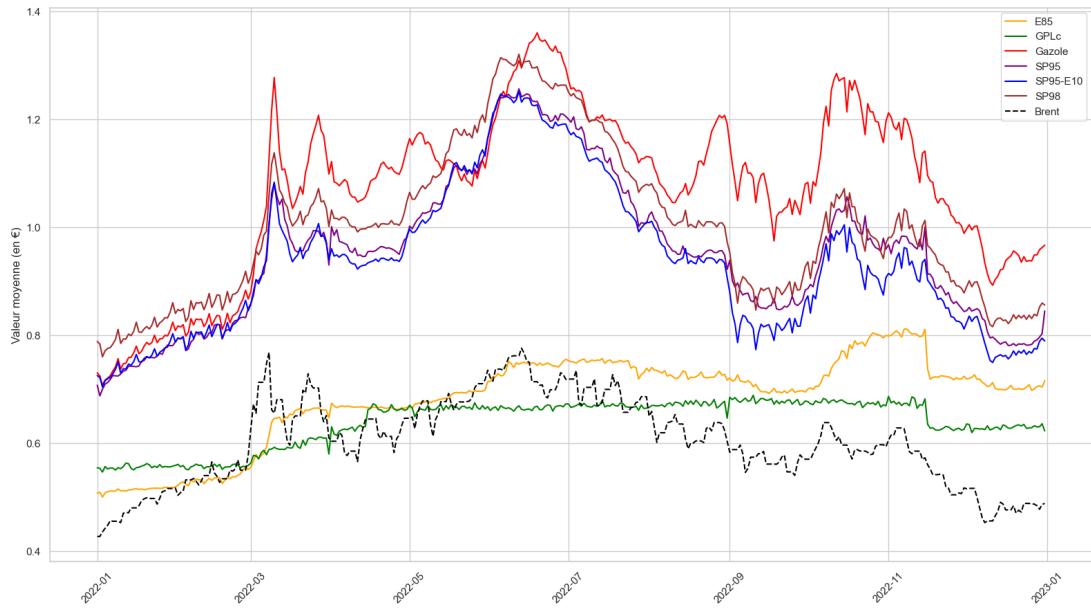
for carburant in df_stats['nom_carburant'].unique():
    subset = df_stats[df_stats['nom_carburant'] == carburant]
    plt.plot(subset['maj'], subset['valeur'], label=carburant,
             color=couleurs_carburants[carburant])

plt.plot(df_brent['maj'],df_brent['litre_brut'],label='Brent', color='black',
         linestyle='--')

plt.ylabel('Valeur moyenne (en €)')
plt.legend()
plt.xticks(rotation=45)
plt.tight_layout()

plt.savefig('graph/Evolution_prix_détail_2022_E85_GPLc_compris.pdf')

plt.show()
```



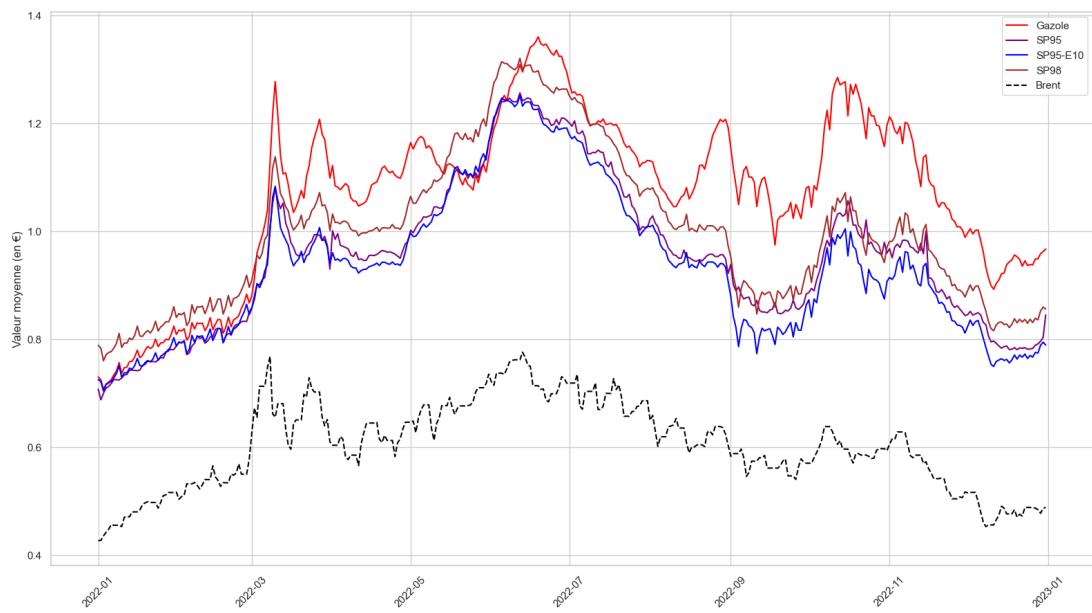
```
[9]: df_stats1=df_stats.loc[~df_stats.iloc[:, 0].isin(['E85', 'GPLc'])]

couleurs_carburants = {
    'Gazole': 'red',
    'SP95': 'purple',
    'SP98': 'brown',
    'SP95-E10': 'blue'
}
for carburant in df_stats1['nom_carburant'].unique():
    subset1 = df_stats1[df_stats1['nom_carburant'] == carburant]
    plt.plot(subset1['maj'], subset1['valeur'], label=carburant, color=couleurs_carburants[carburant])

plt.plot(df_brent['maj'],df_brent['litre_brut'],label='Brent', color='black', linestyle='--')

plt.ylabel('Valeur moyenne (en €)')
plt.legend()
plt.xticks(rotation=45)
plt.tight_layout()

plt.savefig('graph/Evolution_prix_détail_2022.pdf')
plt.show()
```



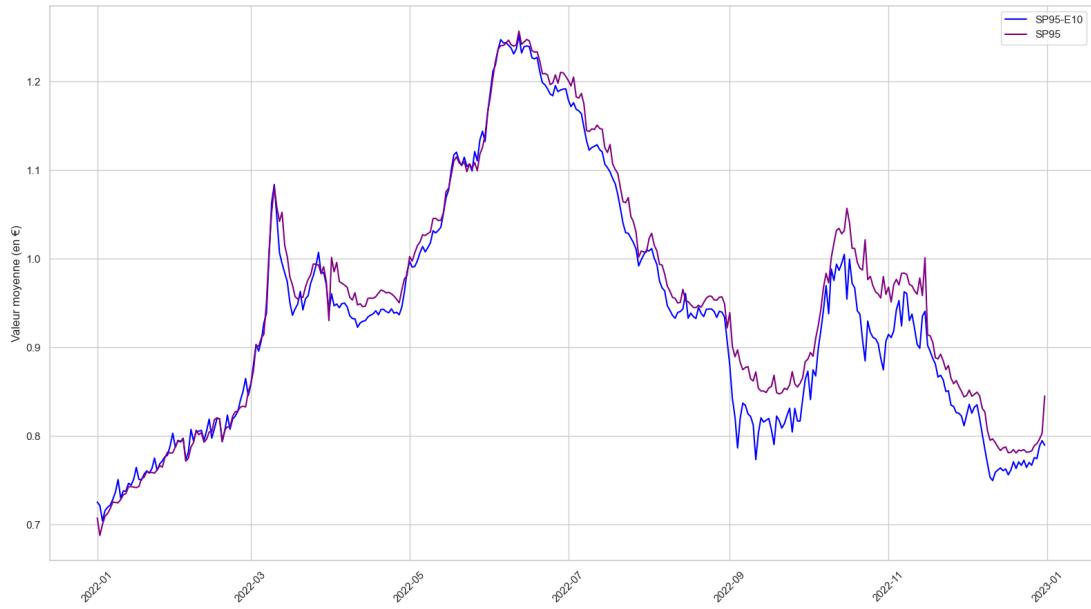
```
[14]: df_stats1=df_stats.loc[~df_stats.iloc[:, 0].isin(['E85', 'GPLc'])]

couleurs_carburants = {
    'Gazole': 'red',
    'SP95': 'purple',
    'SP98': 'brown',
    'SP95-E10': 'blue'
}

subset1 = df_stats1[df_stats1['nom_carburant'] == 'SP95-E10']
subset2 = df_stats1[df_stats1['nom_carburant'] == 'SP95']
plt.plot(subset1['maj'], subset1['valeur'], label='SP95-E10', □
         ↳color=couleurs_carburants['SP95-E10'])
plt.plot(subset1['maj'], subset2['valeur'], label='SP95', □
         ↳color=couleurs_carburants['SP95'])

plt.ylabel('Valeur moyenne (en €)')
plt.legend()
plt.xticks(rotation=45)
plt.tight_layout()

plt.show()
```



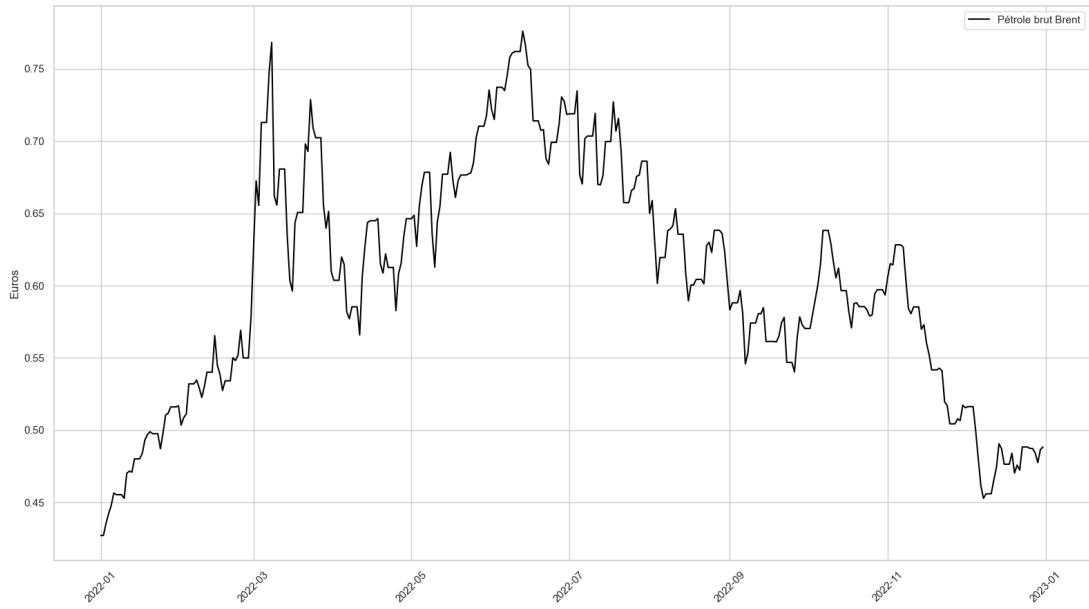
```
[15]: np.corrcoef(subset1.iloc[:, -1], subset2.iloc[:, -1])
```

```
[15]: array([[1.          , 0.98696768],
       [0.98696768, 1.         ]])
```

```
[16]: plt.plot(df_brent['maj'], df_brent['litre_brut'], label='Pétrole brut Brent', color='black')

plt.ylabel('Euros')
plt.xticks(rotation=45)
plt.tight_layout()
plt.legend()
plt.savefig('graph/Evolution_cours_brent_2022_euros.pdf')

plt.show()
```



```
[17]: df_brent['litre_brut'].mean()
```

```
[17]: 0.6035232204907219
```

```
[18]: df_n=df[df['nom_carburant'].isin(['SP95','E10'])]
df_n
```

	cp	id	nom_carburant	id_carburant	maj	valeur	\
138	01	1000001	SP95		2 2022-01-03	0.695633	
139	01	1000001	SP95		2 2022-01-05	0.687300	
140	01	1000001	SP95		2 2022-01-06	0.693967	
141	01	1000001	SP95		2 2022-01-10	0.711467	
142	01	1000001	SP95		2 2022-01-17	0.716467	
...	..	...	...	...	...	...	
4524784	95	95570003	SP95		2 2022-12-30	0.782067	
4526468	95	95600001	SP95		2 2022-01-03	0.723767	
4526469	95	95600001	SP95		2 2022-01-04	0.740433	
4526470	95	95600001	SP95		2 2022-10-28	0.821267	
4532995	95	95712004	SP95		2 2022-04-30	0.918767	
	litre_brut	type	latitude	longitude	ville	\	
138	0.435626	R	46.201	5.198	SAINT-DENIS-LÈS-BOURG		
139	0.447949	R	46.201	5.198	SAINT-DENIS-LÈS-BOURG		
140	0.456551	R	46.201	5.198	SAINT-DENIS-LÈS-BOURG		
141	0.452874	R	46.201	5.198	SAINT-DENIS-LÈS-BOURG		
142	0.484043	R	46.201	5.198	SAINT-DENIS-LÈS-BOURG		
...	...	...	...	...	...		

```

4524784    0.486593    R    49.041      2.342          MOISSELLES
4526468    0.435626    R    48.978      2.291          EAUBONNE
4526469    0.442323    R    48.978      2.291          EAUBONNE
4526470    0.597232    R    48.978      2.291          EAUBONNE
4532995    0.646327    R    49.009      2.537  Roissy-en-France

      stations_rayon_5km  competition_geo
138                      5                  3
139                      5                  3
140                      5                  3
141                      5                  3
142                      5                  3
...
4524784                  8                  3
4526468                  16                 4
4526469                  16                 4
4526470                  16                 4
4532995                  2                  1

```

[327991 rows x 13 columns]

```
[19]: df_ag=pd.DataFrame(df_n.groupby(['maj'])['valeur'].mean()).reset_index()
df_ag
```

```

[19]:      maj      valeur
0  2022-01-01  0.707498
1  2022-01-02  0.687861
2  2022-01-03  0.700118
3  2022-01-04  0.709561
4  2022-01-05  0.712683
...
360 2022-12-27  0.789126
361 2022-12-28  0.791479
362 2022-12-29  0.796477
363 2022-12-30  0.803341
364 2022-12-31  0.845103

```

[365 rows x 2 columns]

```
[20]: np.corrcoef(df_ag['valeur'], df_brent['litre_brut'])
```

```
[20]: array([[1.          , 0.90458718],
       [0.90458718, 1.         ]])
```

```
[21]: np.corrcoef(subset1.iloc[:, -1], df_brent['litre_brut'])
```

```
[21]: array([[1.        , 0.90908428],  
           [0.90908428, 1.        ]])
```

## significativité\_AR\_on\_v5

May 7, 2024

```
[1]: import os
os.chdir("C:/Users/zetru/OneDrive/Mémoire python/")
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
plt.rcParams['figure.figsize'] = 16, 9
plt.rcParams['axes.grid'] = True
import seaborn as sns
sns.set_theme(style="whitegrid")
```

```
[2]: df=pd.read_hdf('df_v5.h5','df_v5')
```

```
[3]: df['nom_carburant'] = df['nom_carburant'].replace({'E10': 'SP95-E10'})
```

```
[4]: df_ar=(pd.DataFrame(df.groupby(['type','nom_carburant','maj'])['valeur'].
mean()))
```

```
[5]: df_ar=df_ar.reset_index()
df_ar
```

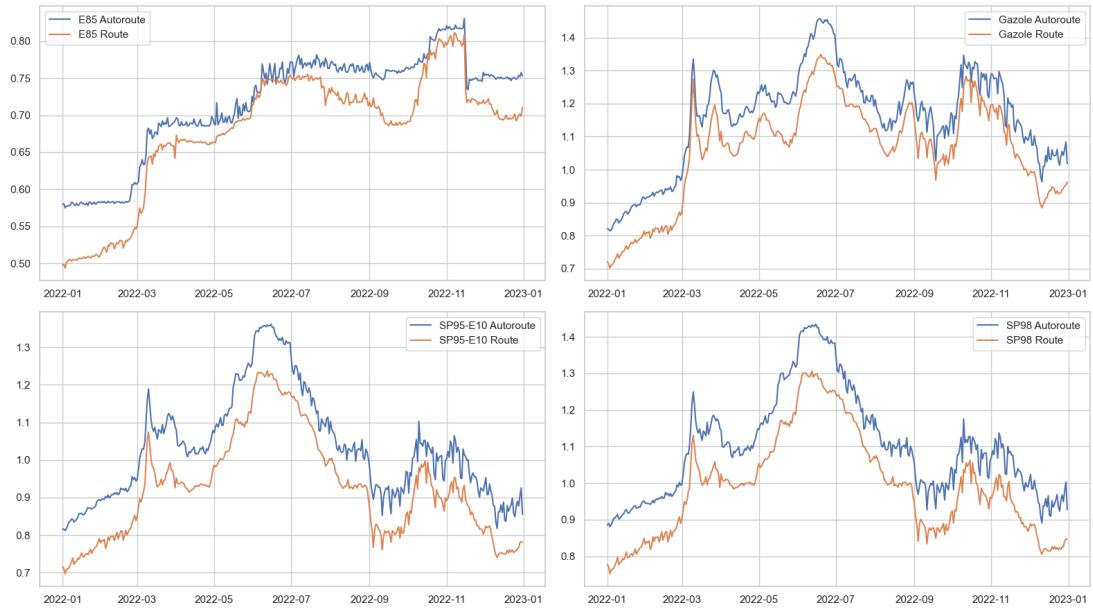
```
[5]:   type nom_carburant      maj    valeur
  0      A             E85 2022-01-01  0.580212
  1      A             E85 2022-01-02  0.580259
  2      A             E85 2022-01-03  0.574908
  3      A             E85 2022-01-04  0.577283
  4      A             E85 2022-01-05  0.577245
...
  ...    ...
  4304     R             SP98 2022-12-27  0.825403
  4305     R             SP98 2022-12-28  0.826612
  4306     R             SP98 2022-12-29  0.841828
  4307     R             SP98 2022-12-30  0.846912
  4308     R             SP98 2022-12-31  0.847749
[4309 rows x 4 columns]
```

```
[7]: df_ar=df_ar[~(df_ar['nom_carburant'].isin(['GPLc','SP95']))]
```

```
[8]: df_ar['type'].unique()
```

```
[8]: array(['A', 'R'], dtype=object)
```

```
[20]: import matplotlib.pyplot as plt
plt.rcParams['axes.grid'] = True
fig, axs = plt.subplots(2, 2, figsize=(16, 9))
n_lignes=2
n_colonnes=3
compte=0
i=0
j=0
for nom in df_ar['nom_carburant'].unique():
    df_mask=df_ar[df_ar['nom_carburant']==nom]
    for type_ar in df_mask['type'].unique():
        ax=axs[i,j]
        if type_ar=='A':
            type_station='Autoroute'
        else:
            type_station='Route'
        df_mask2=df_mask[df_mask['type']==type_ar]
        ax.plot(df_mask2['maj'],df_mask2['valeur'],label=f'{nom} {type_station}')
        ax.legend()
        j+=1
        if j==2:
            j=0
            i+=1
plt.tight_layout()
plt.savefig('graph/comparaison_prix_autoroute.pdf')
plt.show()
```



## stats\_descriptives

May 7, 2024

```
[1]: import os
os.chdir("C:/Users/zetru/OneDrive/Mémoire python/")
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
plt.rcParams['figure.figsize'] = 16, 9
plt.rcParams['axes.grid'] = True
import seaborn as sns
sns.set_theme(style="whitegrid")
```

```
[2]: df=pd.read_hdf('df_v5.h5','df_v5')
```

```
[3]: df['id'].nunique()
```

```
[3]: 9470
```

```
[4]: df['nom_carburant'] = df['nom_carburant'].replace({'E10': 'SP95-E10', 'SP95': 'SP95-E5'})
```

```
[5]: df['type'] = df['type'].replace({'A': 'Autoroute', 'R': 'Route'})
```

```
[6]: df.columns
```

```
[6]: Index(['cp', 'id', 'nom_carburant', 'id_carburant', 'maj', 'valeur',
       'litre_brut', 'type', 'latitude', 'longitude', 'ville',
       'stations_rayon_5km', 'competition_geo'],
       dtype='object')
```

```
[7]: df['maj'] = pd.to_datetime(df['maj'])
df = df.sort_values(by=['id', 'nom_carburant', 'maj'])
df['prix_diff'] = df.groupby(['id', 'nom_carburant'])['maj'].diff().dt.days

resultats = []

carburants = df['nom_carburant'].unique()

for carburant in carburants:
```

```

for type_station in ['Autoroute', 'Route']:
    df_type = df[(df['nom_carburant'] == carburant) & (df['type'] == type_station)]

    if not df_type.empty:
        stats_type = {
            'Carburant': carburant,
            'Type': type_station,
            'Nb obs.': df_type['id'].nunique(),
            'Durée moyenne (en jours)': df_type['prix_diff'].mean(),
            'Ecart-type': df_type['prix_diff'].std(),
            'Q25': df_type['prix_diff'].quantile(0.25),
            'Q50': df_type['prix_diff'].median(),
            'Q75': df_type['prix_diff'].quantile(0.75)
        }
        resultats.append(stats_type)

resultats_df = pd.DataFrame(resultats)

resultats_df = resultats_df[['Carburant', 'Type', 'Nb obs.', 'Durée moyenne (en jours)', 'Ecart-type', 'Q25', 'Q50', 'Q75']]

resultats_df = resultats_df.round(2)

resultats_df

```

[7]:

	Carburant	Type	Nb obs.	Durée moyenne (en jours)	Ecart-type	Q25	Q50	Q75
0	Gazole	Autoroute	436	1.59	1.63	1.0	1.0	1.0
1	Gazole	Route	9023	2.47	3.14	1.0	1.0	1.0
2	SP95-E10	Autoroute	431	1.61	1.77	1.0	1.0	1.0
3	SP95-E10	Route	6808	2.28	2.87	1.0	1.0	1.0
4	SP98	Autoroute	428	1.63	1.90	1.0	1.0	1.0
5	SP98	Route	7701	2.62	3.74	1.0	1.0	1.0
6	SP95-E5	Autoroute	43	3.27	3.02	1.0	1.0	1.0
7	SP95-E5	Route	4317	4.03	5.22	1.0	1.0	1.0
8	E85	Autoroute	161	1.35	2.53	1.0	1.0	1.0
9	E85	Route	3013	2.72	5.57	1.0	1.0	1.0
10	GPLc	Autoroute	319	2.01	6.36	1.0	1.0	1.0
11	GPLc	Route	1129	2.59	8.78	1.0	1.0	1.0

```

6    2.0  4.0
7    2.0  5.0
8    1.0  1.0
9    1.0  2.0
10   1.0  1.0
11   1.0  1.0

[8]: formatters = {col: '{:.2f}'.format for col in resultats_df.
    ↪set_index(['Carburant', 'Type']).select_dtypes(include=['float', 'int']).
    ↪columns}
resultats_df.set_index(['Carburant', 'Type']).to_latex('tex/tableau_majINDEX.
    ↪tex',formatters=formatters)

[9]: df['type'] = df['type'].replace({'Autoroute': 'A', 'Route': 'R'})

[10]: df_a=df[df['type']=='A']
df_r=df[df['type']=='R']

[11]: df_corr=pd.DataFrame(df.
    ↪groupby(['nom_carburant','maj','type'])[['valeur','litre_brut']].mean().
    ↪reset_index()
df_corr_a=pd.DataFrame(df_a.
    ↪groupby(['nom_carburant','maj','type'])[['valeur','litre_brut']].mean().
    ↪reset_index()
df_corr_r=pd.DataFrame(df_r.
    ↪groupby(['nom_carburant','maj','type'])[['valeur','litre_brut']].mean().
    ↪reset_index())

[12]: def f_correlation(x):
    x=x.iloc[:, -2:]
    return x.corr().iloc[0, 1]

[13]: tableau_a=df_corr_a.groupby(['nom_carburant','type']).apply(f_correlation).
    ↪reset_index().rename(columns={0:'Autoroute'}).drop(columns=['type'])
tableau_a

[13]:   nom_carburant  Autoroute
0           E85  0.330473
1          GPLc  0.421298
2         Gazole  0.791620
3  SP95-E10  0.902743
4  SP95-E5  0.854859
5      SP98  0.897102

[14]: tableau_r=df_corr_r.groupby(['nom_carburant','type']).apply(f_correlation).
    ↪reset_index().rename(columns={0:'Route'}).drop(columns=['type'])
tableau_r

```

```
[14]:   nom_carburant      Route
0          E85  0.455217
1          GPLc  0.519583
2          Gazole  0.778601
3    SP95-E10  0.909465
4    SP95-E5  0.904595
5        SP98  0.908698
```

```
[15]: tableau_correlation=pd.concat([tableau_a,tableau_r.iloc[:, -1]],axis=1)
tableau_correlation
```

```
[15]:   nom_carburant  Autoroute      Route
0          E85  0.330473  0.455217
1          GPLc  0.421298  0.519583
2          Gazole  0.791620  0.778601
3    SP95-E10  0.902743  0.909465
4    SP95-E5  0.854859  0.904595
5        SP98  0.897102  0.908698
```

```
[16]: tableau_correlation.to_latex('tex/tableau_corr.tex',index=False)
```

```
[17]: tableau_maj=df.groupby(['id','nom_carburant','type'])['valeur'].size()\
.reset_index()\n    .groupby(['nom_carburant','type'])['valeur']\n    .describe().round(2).reset_index()
tableau_maj
```

```
[17]:   nom_carburant type  count      mean       std     min    25%    50%    75%  \
0          E85     A  161.0  243.81  140.53    1.0  75.00  353.0  360.00
1          E85     R 3013.0  118.55  122.92    1.0  24.00  55.0   207.00
2          GPLc    A  319.0  165.40  139.69    1.0  24.00  131.0  340.00
3          GPLc    R 1129.0  122.49  137.96    1.0  10.00  36.0   251.00
4          Gazole   A  436.0  225.29  105.03   27.0  133.00  202.0  358.25
5          Gazole   R 9023.0  142.56  94.38    1.0  65.00  122.0  197.00
6    SP95-E10    A  431.0  221.58  108.24   20.0  127.00  202.0  358.00
7    SP95-E10    R 6808.0  150.84  96.57    1.0  75.00  131.5  212.00
8    SP95-E5     A   43.0   25.63   49.35    1.0   1.00    1.0   10.50
9    SP95-E5     R 4317.0   75.72   66.93    1.0  31.00   57.0  103.00
10         SP98    A  428.0  218.29  109.48   25.0  123.75  193.0  358.00
11         SP98    R 7701.0  132.65   97.60    1.0  55.00  103.0  191.00

           max
0    364.0
1    365.0
2    364.0
3    365.0
4    365.0
```

```
5  365.0
6  364.0
7  365.0
8  203.0
9  365.0
10 364.0
11 365.0
```

```
[18]: formatters = {col: '{:.2f}'.format for col in tableau_maj.
    ↪select_dtypes(include=['float', 'int']).columns}
tableau_maj.to_latex('tex/tableau_maj.tex', index=False, formatters=formatters)
```

```
[19]: tableau_maj.to_latex('tex/tableau_maj.tex', index=False)
```

# Modèle économétrique

May 7, 2024

```
[1]: import os
os.chdir("C:/Users/zetru/OneDrive/Mémoire python/")
import numpy as np
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt
%matplotlib inline
plt.rcParams['figure.figsize'] = 16, 9
plt.rcParams['axes.grid'] = True
import seaborn as sns
sns.set_theme(style="whitegrid")

[2]: from statsmodels.iolib.summary2 import summary_col
from statsmodels.tsa.stattools import adfuller

[3]: df=pd.read_hdf('df_v5.h5','df_v5')

[135]: df=df[(df['nom_carburant']!='SP95')]
df.replace('E10', 'SP95E10', inplace=True)
df['ln_PP']=np.log(df['valeur'])
df['ln_Brent']=np.log(df['litre_brut'])
df['competition_geo']=df['competition_geo'].astype(float)

[136]: df['nom_carburant'].unique()

[136]: array(['Gazole', 'SP98', 'SP95E10', 'GPLc', 'E85'], dtype=object)

[137]: df_Gazole_A=df[(df['nom_carburant']=='Gazole')&(df['type']=='A')]
df_Gazole_R=df[(df['nom_carburant']=='Gazole')&(df['type']=='R')]
df_SP95E10_A=df[(df['nom_carburant']=='SP95E10')&(df['type']=='A')]
df_SP95E10_R=df[(df['nom_carburant']=='SP95E10')&(df['type']=='R')]
df_SP98_A=df[(df['nom_carburant']=='SP98')&(df['type']=='A')]
df_SP98_R=df[(df['nom_carburant']=='SP98')&(df['type']=='R')]
```

### 0.0.1 Relation de long-terme (Equation 1)

```
[141]: # Liste des DataFrames pour autoroutes et routes
dfs = [df_Gazole_A, df_Gazole_R, df_SP95E10_A, df_SP95E10_R, df_SP98_A, df_SP98_R]
df_autoroutes = [df_Gazole_A, df_SP95E10_A, df_SP98_A]
df_routes = [df_Gazole_R, df_SP95E10_R, df_SP98_R]
labels = ['Gazole', 'SP95E10', 'SP98']

# Fonction pour effectuer la régression linéaire
def run_regression_LT(df):
    y = df['ln_PP'] # prix à la pompe HT
    X = df[['ln_Brent']] # prix du Brent
    dummies = pd.get_dummies(df['competition_geo'], drop_first=True, prefix='comp_geo', dtype=float)

    X = pd.concat([X, dummies], axis=1)

    X = sm.add_constant(X)
    model = sm.OLS(y, X)
    results = model.fit()
    df['residus'] = results.resid # stocker les résidus du modèle
    return results

results_autoroutes = [run_regression_LT(df) for df in df_autoroutes]
results_routes = [run_regression_LT(df) for df in df_routes]

summary_table_autoroutes = summary_col(results_autoroutes, stars=True, model_names=labels,
                                         info_dict={'No. observations': lambda x: f'{int(x.nobs)}'})
summary_table_routes = summary_col(results_routes, stars=True, model_names=labels,
                                    info_dict={'No. observations': lambda x: f'{int(x.nobs)}'})
```

C:\Users\zetru\AppData\Local\Temp\ipykernel\_50264\71725348.py:18:

SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

df['residus'] = results.resid # stocker les résidus du modèle

C:\Users\zetru\AppData\Local\Temp\ipykernel\_50264\71725348.py:18:

SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

```
    df['residus'] = results.resid # stocker les résidus du modèle
C:\Users\zetru\AppData\Local\Temp\ipykernel_50264\71725348.py:18:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

```
    df['residus'] = results.resid # stocker les résidus du modèle
C:\Users\zetru\AppData\Local\Temp\ipykernel_50264\71725348.py:18:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

```
    df['residus'] = results.resid # stocker les résidus du modèle
C:\Users\zetru\AppData\Local\Temp\ipykernel_50264\71725348.py:18:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

```
    df['residus'] = results.resid # stocker les résidus du modèle
C:\Users\zetru\AppData\Local\Temp\ipykernel_50264\71725348.py:18:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

```
    df['residus'] = results.resid # stocker les résidus du modèle
```

```
[105]: print("Résultats pour les routes:")
display(summary_table_routes)
print("\nRésultats pour les autoroutes:")
display(summary_table_autoroutes)
```

Résultats pour les routes:

Standard errors in parentheses.

\* p<.1, \*\* p<.05, \*\*\*p<.01

	Gazole	SP95E10	SP98
const	0.4893*** (0.0004)	0.4004*** (0.0004)	0.4455*** (0.0003)
ln_Brent	0.8657*** (0.0007)	0.9625*** (0.0006)	0.9203*** (0.0006)
comp_geo_2.0	-0.0027*** (0.0003)	0.0000 (0.0003)	-0.0009*** (0.0003)
comp_geo_3.0	-0.0044*** (0.0003)	-0.0006*** (0.0002)	0.0004 (0.0002)
comp_geo_4.0	0.0132*** (0.0003)	0.0204*** (0.0002)	0.0241*** (0.0002)
R-squared	0.5396	0.7008	0.6861
R-squared Adj.	0.5396	0.7008	0.6861
No. observations	1286334	1026934	1021550

	Gazole	SP95E10	SP98
const	0.5445*** (0.0013)	0.4868*** (0.0013)	0.5267*** (0.0013)
ln_Brent	0.7578*** (0.0024)	0.8642*** (0.0024)	0.8161*** (0.0023)
comp_geo_2.0	-0.0227*** (0.0008)	-0.0265*** (0.0008)	-0.0269*** (0.0008)
comp_geo_3.0	-0.0360*** (0.0010)	-0.0455*** (0.0010)	-0.0423*** (0.0009)
comp_geo_4.0	-0.0617*** (0.0012)	-0.0784*** (0.0012)	-0.0751*** (0.0011)
R-squared	0.5034	0.5899	0.5774
R-squared Adj.	0.5034	0.5899	0.5774
No. observations	98226	95499	93426

Résultats pour les autoroutes:

Standard errors in parentheses.

\* p<.1, \*\* p<.05, \*\*\*p<.01

```
[9]: def export_to_latex(summary, filename):
    try:
        latex_content = summary.as_latex()
        with open(filename, 'w') as file:
            file.write(latex_content)
        print(f"Le fichier {filename} a été créé avec succès.")
    except Exception as e:
        print(f"Une erreur est survenue lors de l'écriture du fichier LaTeX : {e}")
```

```
[10]: export_to_latex(summary_table_autoroutes, 'tex/LT_A.tex')
export_to_latex(summary_table_routes, 'tex/LT_R.tex')
```

Le fichier tex/LT\_A.tex a été créé avec succès.  
Le fichier tex/LT\_R.tex a été créé avec succès.

### 0.0.2 Test ADF sur les résidus de la relation de long-terme

```
[11]: results_adf = [adfuller(df['residus'], maxlag=0) for df in dfs]
results_adf

[12]: labels_adf = ['Gazole', 'Gazole', 'SP95E10', 'SP95E10', 'SP98', 'SP98']
typelabels = ['Autoroute', 'Route', 'Autoroute', 'Route', 'Autoroute', 'Route']

adf_stats = {
    'Carburant': labels_adf,
    'Type': typelabels,
    'ADF Statistic': [result[0] for result in results_adf],
    'p-value': [result[1] for result in results_adf],
    'Used Lag': [result[2] for result in results_adf],
    'Obs.': [result[3] for result in results_adf]
}

adf_results_df = pd.DataFrame(adf_stats)
adf_results_df
```

	Carburant	Type	ADF Statistic	p-value	Used Lag	Obs.
0	Gazole	Autoroute	-58.633813	0.0	0	98225
1	Gazole	Route	-264.592493	0.0	0	1286333
2	SP95E10	Autoroute	-58.246941	0.0	0	95498
3	SP95E10	Route	-275.567863	0.0	0	1026933
4	SP98	Autoroute	-56.296158	0.0	0	93425
5	SP98	Route	-280.174678	0.0	0	1021549

```
[13]: adf_results_df.to_latex('tex/adf_results.tex', index=False)
```

### 0.0.3 Relation de court-terme : Modèle à correction d'erreur sans asymétrie (Equation 2)

```
[14]: labels = ['Gazole', 'SP95E10', 'SP98']
# Préparation des données pour la régression de correction d'erreur à court
# terme
def prepare_data_for_ecm(df):
    df = df.copy()

    df['id_index'] = df.groupby('id').ngroup()

    df.loc[:, 'delta_ln_PP'] = df.groupby('id')['ln_PP'].transform(lambda x: x.
    -diff())
```

```

    df.loc[:, 'delta_ln_Brent'] = df.groupby('id')['ln_Brent'].transform(lambda x: x.diff())
    df.loc[:, 'error_correction'] = df.groupby('id')['residus'].transform(lambda x: x.shift(1))

    df.loc[:, 'delta_ln_PP_1'] = df.groupby('id')['delta_ln_PP'].transform(lambda x: x.shift(1))
    df.loc[:, 'delta_ln_Brent_1'] = df.groupby('id')['delta_ln_Brent'].transform(lambda x: x.shift(1))
    df.loc[:, 'delta_ln_Brent_2'] = df.groupby('id')['delta_ln_Brent_1'].transform(lambda x: x.shift(1))

    df.dropna(subset=['delta_ln_PP', 'delta_ln_Brent', 'error_correction',
                      'delta_ln_PP_1', 'delta_ln_Brent_1', 'delta_ln_Brent_2'], inplace=True)
    df.drop(columns='id_index', inplace=True)

    return df

# Fonction pour effectuer la régression de correction d'erreur à court terme
def run_ecm(df):
    df = prepare_data_for_ecm(df)
    X_ecm = df[['delta_ln_Brent', 'delta_ln_Brent_1', 'delta_ln_Brent_2', 'delta_ln_PP_1', 'error_correction']]
    y_ecm = df['delta_ln_PP']

    model_ecm = sm.OLS(y_ecm, X_ecm)
    results_ecm = model_ecm.fit()
    return results_ecm

results_autoroutes_ecm = [run_ecm(df) for df in df_autoroutes]
results_routes_ecm = [run_ecm(df) for df in df_routes]
summary_table_autoroutes_ecm = summary_col(results_autoroutes_ecm, stars=True,
                                             model_names=labels,
                                             info_dict={'No. observations': lambda x: f'{int(x.nobs)}'})
summary_table_routes_ecm = summary_col(results_routes_ecm, stars=True,
                                         model_names=labels,
                                         info_dict={'No. observations': lambda x: f'{int(x.nobs)}'})

```

```
[15]: print("Résultats ECM pour les routes:")
display(summary_table_routes_ecm)
print("\nRésultats ECM pour les autoroutes:")
display(summary_table_autoroutes_ecm)
```

	Gazole	SP95E10	SP98
delta_ln_Brent	0.1697*** (0.0008)	0.1523*** (0.0008)	0.1748*** (0.0008)
delta_ln_Brent_1	0.2162*** (0.0009)	0.1756*** (0.0009)	0.1729*** (0.0008)
delta_ln_Brent_2	0.1706*** (0.0009)	0.1482*** (0.0009)	0.1372*** (0.0008)
delta_ln_PP_1	0.0074*** (0.0008)	-0.0339*** (0.0010)	-0.0508*** (0.0010)
error_correction	-0.0752*** (0.0003)	-0.0753*** (0.0004)	-0.0817*** (0.0004)
R-squared	0.2102	0.1614	0.1727
R-squared Adj.	0.2102	0.1614	0.1726
No. observations	1259283	1006755	998503

	Gazole	SP95E10	SP98
delta_ln_Brent	0.0644*** (0.0027)	0.0887*** (0.0026)	0.0846*** (0.0024)
delta_ln_Brent_1	0.1638*** (0.0028)	0.1566*** (0.0026)	0.1431*** (0.0025)
delta_ln_Brent_2	0.1682*** (0.0028)	0.1692*** (0.0027)	0.1570*** (0.0025)
delta_ln_PP_1	0.0112*** (0.0031)	-0.0933*** (0.0032)	-0.0806*** (0.0032)
error_correction	-0.0486*** (0.0009)	-0.0326*** (0.0008)	-0.0321*** (0.0008)
R-squared	0.1420	0.1152	0.1144
R-squared Adj.	0.1420	0.1151	0.1144
No. observations	96918	94206	92142

Résultats ECM pour les routes:

Standard errors in parentheses.

\* p<.1, \*\* p<.05, \*\*\*p<.01

Résultats ECM pour les autoroutes:

Standard errors in parentheses.

\* p<.1, \*\* p<.05, \*\*\*p<.01

```
[16]: export_to_latex(summary_table_autoroutes_ecm, 'tex/ECM_A.tex')
export_to_latex(summary_table_routes_ecm, 'tex/ECM_R.tex')
```

Le fichier tex/ECM\_A.tex a été créé avec succès.

Le fichier tex/ECM\_R.tex a été créé avec succès.

#### 0.0.4 Relation de court-terme : Modèle à correction d'erreur avec asymétrie (Equation 3)

labels = ['Gazole', 'SP95E10', 'SP98']

### 1 Préparation des données pour la régression de correction d'erreur à court terme avec asymétrie

```
def prepare_data_for_ecm_asym(df):    df['delta_ln_PP'] = df.groupby('id')['ln_PP'].diff()    df['delta_ln_Brent'] = df.groupby('id')['ln_Brent'].diff()    df['error_correction'] = df.groupby('id')['residus'].shift(1)

df['delta_ln_PP_1'] = df.groupby('id')['delta_ln_PP'].shift(1)
df['delta_ln_Brent_1'] = df.groupby('id')['delta_ln_Brent'].shift(1)
df['delta_ln_Brent_2'] = df.groupby('id')['delta_ln_Brent_1'].shift(1)

# Créer les variables indicatrices pour les ajustements asymétriques
df['lambda'] = (df['delta_ln_Brent_1'] > 0).astype(int)
df['error_correction_pos'] = df['lambda'] * df['error_correction']
df['error_correction_neg'] = (1 - df['lambda']) * df['error_correction']

df.dropna(subset=['delta_ln_PP', 'delta_ln_Brent', 'error_correction',
                  'delta_ln_PP_1', 'delta_ln_Brent_1', 'delta_ln_Brent_2',
                  'error_correction_pos', 'error_correction_neg'], inplace=True)

return df
```

### 2 Fonction pour effectuer la régression de correction d'erreur à court terme avec asymétrie

```
def run_ecm_asym(df): df = prepare_data_for_ecm_asym(df)

X_ecm = df[['delta_ln_Brent', 'delta_ln_Brent_1', 'delta_ln_Brent_2', 'delta_ln_PP_1',
            'error_correction_pos', 'error_correction_neg']]
y_ecm = df['delta_ln_PP']
model_ecm = sm.OLS(y_ecm, X_ecm)
results_ecm = model_ecm.fit()
return results_ecm

results_autoroutes_ecm_asym = [run_ecm_asym(df) for df in df_autoroutes]
results_routes_ecm_asym = [run_ecm_asym(df) for df in df_routes]

summary_table_autoroutes_ecm_asym = summary_col(results_autoroutes_ecm_asym,
                                                 stars=True, model_names=labels, info_dict={'No. observations': lambda x: f'{int(x.nobs)}'})
summary_table_routes_ecm_asym = summary_col(results_routes_ecm_asym, stars=True,
                                             model_names=labels, info_dict={'No. observations': lambda x: f'{int(x.nobs)}'})
```

	Gazole	SP95E10	SP98
delta_ln_Brent	0.1698*** (0.0008)	0.1524*** (0.0008)	0.1749*** (0.0008)
delta_ln_Brent_1	0.2164*** (0.0009)	0.1750*** (0.0009)	0.1726*** (0.0008)
delta_ln_Brent_2	0.1705*** (0.0009)	0.1483*** (0.0009)	0.1372*** (0.0008)
delta_ln_PP_1	0.0057*** (0.0009)	-0.0344*** (0.0010)	-0.0511*** (0.0010)
error_correction_pos	-0.0857*** (0.0004)	-0.0818*** (0.0005)	-0.0854*** (0.0005)
error_correction_neg	-0.0646*** (0.0004)	-0.0695*** (0.0005)	-0.0781*** (0.0005)
R-squared	0.2111	0.1617	0.1727
R-squared Adj.	0.2111	0.1617	0.1727
No. observations	1259283	1006755	998503

```
[147]: print("Résultats ECM avec asymétrie pour les routes:")
display(summary_table_routes_ecm_asym)
print("\nRésultats ECM avec asymétrie pour les autoroutes:")
display(summary_table_autoroutes_ecm_asym)
```

Résultats ECM avec asymétrie pour les routes:

Standard errors in parentheses.

\* p<.1, \*\* p<.05, \*\*\*p<.01

Résultats ECM avec asymétrie pour les autoroutes:

Standard errors in parentheses.

\* p<.1, \*\* p<.05, \*\*\*p<.01

```
[19]: export_to_latex(summary_table_autoroutes_ecm_asym, 'tex/ECM_asym_A.tex')
export_to_latex(summary_table_routes_ecm_asym, 'tex/ECM_asym_R.tex')
```

Le fichier tex/ECM\_asym\_A.tex a été créé avec succès.

Le fichier tex/ECM\_asym\_R.tex a été créé avec succès.

```
[20]: hypothesis='error_correction_pos = error_correction_neg'
test_results_r = []
for results in results_autoroutes_ecm_asym:
    test = results.t_test(hypothesis)
    test_results_r.append(test.pvalue.item())

test_results_a = []
for results in results_routes_ecm_asym:
    test = results.t_test(hypothesis)
```

	Gazole	SP95E10	SP98
delta_ln_Brent	0.0651*** (0.0027)	0.0887*** (0.0026)	0.0847*** (0.0024)
delta_ln_Brent_1	0.1625*** (0.0028)	0.1560*** (0.0026)	0.1424*** (0.0025)
delta_ln_Brent_2	0.1683*** (0.0028)	0.1692*** (0.0027)	0.1570*** (0.0025)
delta_ln_PP_1	0.0098*** (0.0031)	-0.0930*** (0.0032)	-0.0804*** (0.0032)
error_correction_pos	-0.0619*** (0.0012)	-0.0392*** (0.0013)	-0.0382*** (0.0012)
error_correction_neg	-0.0369*** (0.0012)	-0.0275*** (0.0011)	-0.0274*** (0.0011)
R-squared	0.1441	0.1157	0.1149
R-squared Adj.	0.1441	0.1156	0.1148
No. observations	96918	94206	92142

```

test_results_a.append(test.pvalue.item())

alpha = 0.05

wald_autoroutes = pd.DataFrame({
    'Carburant': labels,
    'P-valeur': test_results_r,
    'Interprétation': ['$\delta_3 \neq \delta_4$' if p < alpha else
        '$\delta_3=\delta_4$' for p in test_results_r]
})

wald_routes = pd.DataFrame({
    'Carburant': labels,
    'P-valeur': test_results_a,
    'Interprétation': ['$\delta_3 \neq \delta_4$' if p < alpha else
        '$\delta_3=\delta_4$' for p in test_results_a]
})

```

```
[21]: print("Résultats de test pour routes:")
display(wald_routes)
print("\nRésultats de test pour autoroutes:")
display(wald_autoroutes)
```

Résultats de test pour routes:

	Carburant	P-valeur	Interprétation
0	Gazole	0.000000e+00	$\delta_3 \neq \delta_4$
1	SP95E10	4.197443e-70	$\delta_3 \neq \delta_4$
2	SP98	1.568377e-23	$\delta_3 \neq \delta_4$

Résultats de test pour autoroutes:

	Carburant	P-valeur	Interprétation
0	Gazole	3.598368e-53	$\delta_3 \neq \delta_4$
1	SP95E10	1.139570e-12	$\delta_3 \neq \delta_4$
2	SP98	1.812037e-11	$\delta_3 \neq \delta_4$

```
[22]: def prepare_wald_dataframe(results_routes, results_autoroutes, hypothesis, alpha=0.05, labels=None):
    # Calcul des p-valeurs pour chaque ensemble de résultats et création des données pour le DataFrame
    data = []
    for label, results_r, results_a in zip(labels, results_routes, results_autoroutes):
        for result, road_type in zip([results_r, results_a], ['Route', 'Autoroute']):
            pvalue = result.t_test(hypothesis).pvalue.item()
            interpretation = '$\theta^+ \neq \theta^-' if pvalue < alpha
            else '$\theta^+ = \theta^-'
            data.append([label, road_type, pvalue, interpretation])
    return pd.DataFrame(data, columns=['Carburant', 'Type', 'P-valeur', 'Interprétation'])

hypothesis = 'error_correction_pos = error_correction_neg'

wald_df = prepare_wald_dataframe(results_routes_ecm_asym, results_autoroutes_ecm_asym, hypothesis, alpha=0.05, labels=labels)
wald_df
```

```
[22]:   Carburant      Type      P-valeur      Interprétation
0   Gazole       Route  0.000000e+00  $\theta^+ \neq \theta^-$
1   Gazole     Autoroute  3.598368e-53  $\theta^+ \neq \theta^-$
2   SP95E10      Route  4.197443e-70  $\theta^+ \neq \theta^-$
3   SP95E10    Autoroute  1.139570e-12  $\theta^+ \neq \theta^-$
4       SP98       Route  1.568377e-23  $\theta^+ \neq \theta^-"
5       SP98     Autoroute  1.812037e-11  $\theta^+ \neq \theta^-"
```

```
[23]: def prepare_wald(results_routes, results_autoroutes, hypothesis, alpha=0.05, labels=None):
    # Calcul des p-valeurs pour chaque ensemble de résultats et création des données pour le DataFrame
    index = []
    data = []
    for label, results_r, results_a in zip(labels, results_routes, results_autoroutes):
```

```

        for result, road_type in zip([results_r, results_a], ['Route', u
↪'Autoroute']):
            pvalue = result.t_test(hypothesis).pvalue.item()
            interpretation = '$\\theta^+ \\neq \\theta^-$' if pvalue < alpha_u
↪else '$\\theta^+ = \\theta^-$'
            index.append((label, road_type))
            data.append({'P-valeur': pvalue, 'Interprétation': interpretation})

    multi_index = pd.MultiIndex.from_tuples(index, names=['Carburant', 'Type'])
    return pd.DataFrame(data, index=multi_index)

hypothesis = 'error_correction_pos = error_correction_neg'

wald_df = prepare_wald(results_routes_ecm_asym, results_autoroutes_ecm_asym, u
↪hypothesis, alpha=0.05, labels=labels)

wald_df

```

[23]:

		P-valeur	Interprétation
Carburant	Type		
	Route	0.000000e+00	$\theta^+ \neq \theta^-$
Gazole	Autoroute	3.598368e-53	$\theta^+ \neq \theta^-$
	Route	4.197443e-70	$\theta^+ \neq \theta^-$
SP95E10	Autoroute	1.139570e-12	$\theta^+ \neq \theta^-$
	Route	1.568377e-23	$\theta^+ \neq \theta^-$
SP98	Autoroute	1.812037e-11	$\theta^+ \neq \theta^-$
	Route		

[24]: wald\_df.to\_latex('tex/wald\_df.tex')

wald\_routes.to\_latex('tex/wald\_R.tex',index=False) wald\_autoroutes.to\_latex('tex/wald\_A.tex',index=False)