

17 / 20

```
from pathlib import Path
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
import seaborn as sns

# Lire fichier et Filtrage -----
df = pd.read_csv(
    Path("data/CRSP_dsf_v2.csv.gz"),
    parse_dates=["DlyCaldt"],
    usecols=["Ticker", "DlyCaldt", "DlyRet", "DlyVol"],
)
tickers = ["AMZN", "BA", "IBM"]
df = df[df["Ticker"].isin(tickers)].sort_values(["Ticker", "DlyCaldt"])
print(df)
```

	Ticker	DlyCaldt	DlyRet	DlyVol
2	AMZN	2023-01-03	0.021667	76643433
5	AMZN	2023-01-04	-0.007924	68762793
8	AMZN	2023-01-05	-0.023726	67855686
11	AMZN	2023-01-06	0.035611	83157846
14	AMZN	2023-01-09	0.014870	65166682
...
735	IBM	2023-12-22	0.008459	2442715
738	IBM	2023-12-26	0.006599	1772443
741	IBM	2023-12-27	0.001532	3234550
744	IBM	2023-12-28	0.001774	2071313
747	IBM	2023-12-29	-0.001221	2526169

[750 rows x 4 columns]

Vérification des données.

Statistiques descriptives?

Lecture

Le fichier CRSP est lu en lisant que les colonnes sélectionnées, dont Ticker, DlyCaldt, DlyRet et DlyVol, tous en convertissant les colonnes de dates en format datetime.

Filtrage

Le DataFrame va ensuite sélectionner les lignes correspondant au ticker AMZN, BA et IBM. Finalement, elle est triée par Ticker et DlyCaldt.

Difficulté

Nous avons dû regarder le contenu du fichier pour pouvoir bien sélectionner les noms des colonnes qui correspond à ceux qu'on doit utiliser selon le mandat du

travail, ce qui peut être mélangeant.

```
# Rendement à numérique
df["DlyRet"] = pd.to_numeric(df["DlyRet"], errors="coerce")
df = df.dropna(subset=["DlyRet", "DlyVol"])
print(df)
```

	Ticker	DlyCalDt	DlyRet	DlyVol
2	AMZN	2023-01-03	0.021667	76643433
5	AMZN	2023-01-04	-0.007924	68762793
8	AMZN	2023-01-05	-0.023726	67855686
11	AMZN	2023-01-06	0.035611	83157846
14	AMZN	2023-01-09	0.014870	65166682
...
735	IBM	2023-12-22	0.008459	2442715
738	IBM	2023-12-26	0.006599	1772443
741	IBM	2023-12-27	0.001532	3234550
744	IBM	2023-12-28	0.001774	2071313
747	IBM	2023-12-29	-0.001221	2526169

[750 rows x 4 columns]

Conversion

Nous avons vérifié la conversion des données de la colonne DlyRet en numérique pour permettre de convertir les valeurs impossibles à convertir en NaN

Nettoyage

Avec la fonction Dropna, nous allons supprimer les valeurs manquantes dans la colonne DlyRet DlyVol.

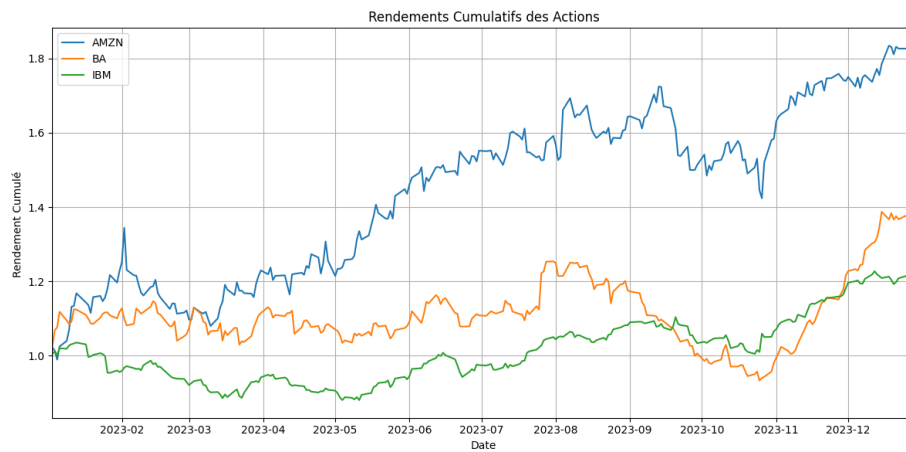
```
# 1. Rendement cumulé -----
df["Rendement_Cumulatif"] = df.groupby("Ticker")["DlyRet"].transform(
    lambda x: (1 + x).cumprod()
)
plt.figure(figsize=(12, 6))

for ticker in tickers:
    ticker_data = df[df["Ticker"] == ticker]
    plt.plot(ticker_data["DlyCalDt"], ticker_data["Rendement_Cumulatif"], label=ticker)

min_date = df["DlyCalDt"].min()
max_date = df["DlyCalDt"].max()

plt.xlim(min_date, max_date)
```

```
plt.title("Rendements Cumulatifs des Actions")
plt.xlabel("Date")
plt.ylabel("Rendement Cumulé ")
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.savefig("rendements_cumulatifs.png", dpi=300)
plt.show()
```



Calcul du rendement cumulatif

Pour calculer le rendement cumulatif, nous allons utiliser la formule $(y) = (1+r_1) \times (1+r_2) \times (1+r_t)$ pour obtenir le rendement cumulé à une date précise. Avant tout, nous avons utilisé la fonction `groupby('Ticker')` pour séparer les données selon Ticker afin de calculer leur rendement cumulé séparément. Ensuite, `['DlyRet']` nous permettra d'utiliser la colonne des rendements quotidiens.

Enfin nous avons utilisé la fonction `transform(lambda x: (1 + x).cumprod())` pour calculer le produit cumulatif pour chaque Ticker en ajoutant 1 à chaque valeur de `['DlyRet']`.

Analyse du graphique

Dans le graphique ci-dessus, on peut voir que l'action Amazon a le rendement cumulatif le plus élevé, suivi de Boeing et IBM. Après avoir été semblable avec les autres actions, le rendement cumulatif d'Amazon a commencé à augmenter plus rapidement à partir du mois d'avril 2023. Alors, sur la période étudiée (2023), l'action Amazon est la plus performante avec 80% de rendement cumulatif, suivi de Boeing avec près de 40% et de IBM avec 20%.

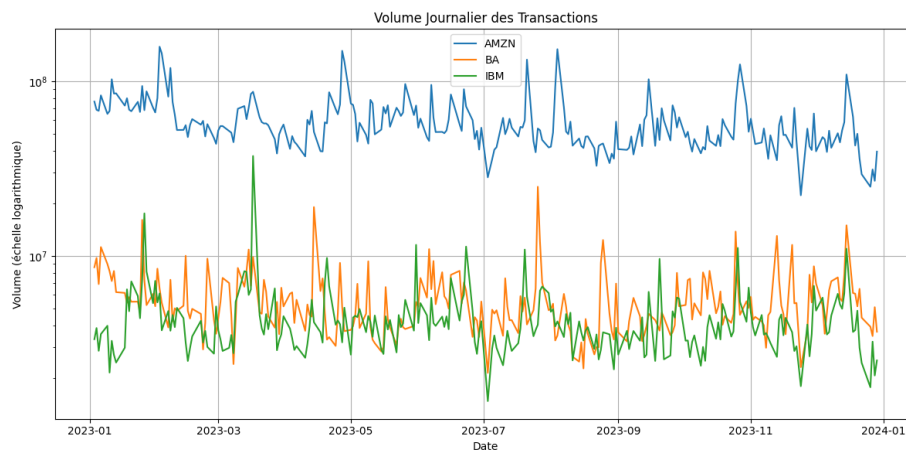
1.5/2

Difficulté,

Un problème récurrent est qu'on voyait seulement un des rendements cumulatifs dans le tableau, pour ce faire, nous avons fait une itération sur chaque Ticker qui va permettre d'appliquer la fonction à chacun.

```
# Volumes de transaction
plt.figure(figsize=(12, 6))
for ticker in tickers:
    data = df[df["Ticker"] == ticker]
    plt.plot(data["DlyCalDt"], data["DlyVol"], label=ticker)

plt.title("Volume Journalier des Transactions")
plt.xlabel("Date")
plt.ylabel("Volume (échelle logarithmique)")
plt.yscale("log")
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.savefig("volumes.png", dpi=300)
plt.show()
```



Itération Non seulement nous avons créé une nouvelle figure, nous avons fait une itération pour permettre de filtrer le Dataframe de chaque action Pour chaque action dans Ticker afin SEULEMENT les tickers importants , dont AMZN ,BA ,IBM.

Analyse des résultats Dans le graphique ci-dessous, on peut voir que le volume des transactions de l'action Amazon est beaucoup plus élevé que celui des actions Boeing et d'IBM, qui eux sont comparables. Le volume des transactions de l'action Amazon est en moyenne de plus près de 10^8 , alors que celui des actions Boeing et IBM est en moyenne un peu sous 10^7 . On peut également voir

la présence d'une certaine volatilité de la part des trois actions, par la présence de mouvements vers le haut et le bas de manière répétée et significative.

Test de normalité -----

```
def plot_distribution(ticker, data):
    fig, axes = plt.subplots(1, 2, figsize=(12, 4))

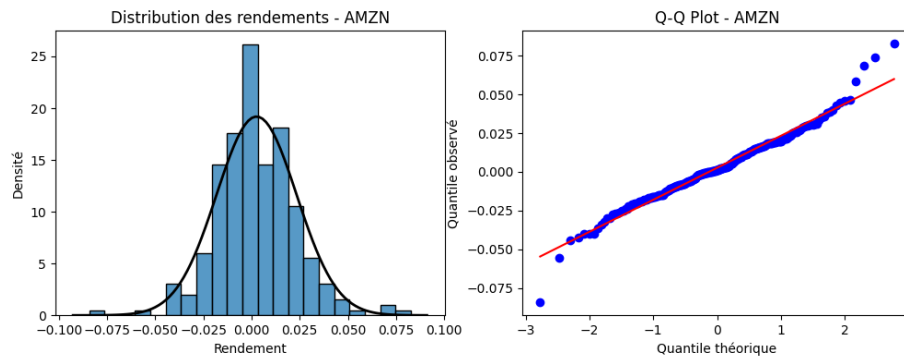
    # Histogramme avec distribution normale

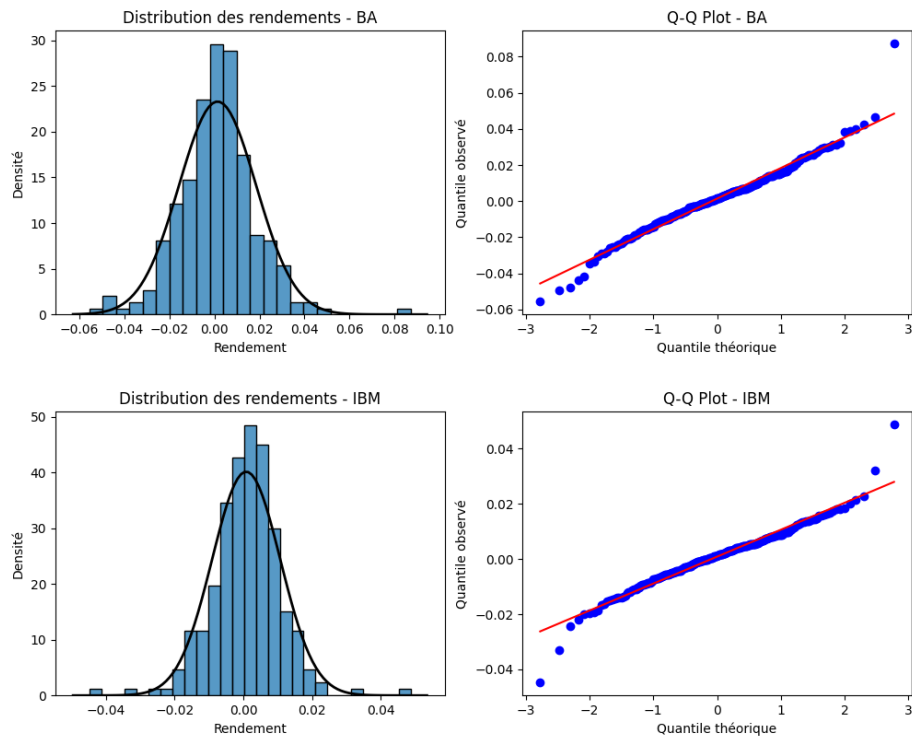
    sns.histplot(data, kde=False, ax=axes[0], stat="density")
    xmin, xmax = axes[0].get_xlim()
    x = np.linspace(xmin, xmax, 100)
    p = stats.norm.pdf(x, data.mean(), data.std())
    axes[0].plot(x, p, "k", linewidth=2)
    axes[0].set_title(f"Distribution des rendements - {ticker}")
    axes[0].set_xlabel("Rendement")
    axes[0].set_ylabel("Densité")

    # Q-Q Plot
    stats.probplot(data, dist="norm", plot=axes[1])
    axes[1].set_title(f"Q-Q Plot - {ticker}")
    axes[1].set_xlabel("Quantile théorique")
    axes[1].set_ylabel("Quantile observé")
    plt.savefig(f"normalite_{ticker}.png", dpi=300)
    plt.show()

for ticker in tickers:
    returns = df[df["Ticker"] == ticker]["DlyRet"]

    plot_distribution(ticker, returns)
```





Définir création de figure

la fonction possède deux paramètres importants, dont `ticker` et `data`(rendement à analyser). le subplot va permettre de créer une figure avec 2 sous-graphiques

Densité

Histogramme

Avec la fonction `sns` nous avons pu créer un histogramme à l'aide de l'utilisation de plusieurs paramètres ci-dessus.

Q-Q plot

Nous avons utilisé la fonction `stats.probplot` pour faire un graphique Q-Q plot en comparant les quantiles théoriques à ceux qui sont observés.

Difficulté

Un problème rencontré est l'affichage de seulement des graphiques pour 1 ticker. Pour ce faire, nous avons mis une itération `'for ticker in tickers'` en filtrant le

dataframe pour sélectionner que les lignes où la colonne Ticker correspond au ticker actuel.

Enfin, nous avons appelé la fonction en appelant ticker et return pour permettre de visualiser les histogrammes et Q-Q Plot.

Analyse de l'histogramme du Q-Q Plot

En regardant les graphiques de distribution et de Q-Q Plot, on peut voir que les rendements des actions Amazon, Boeing et IBM ne suivent pas une distribution normale parfaite, particulièrement pour les valeurs extrêmes au début et à la fin de la distribution dans les graphique Q-Q Plot. Bien que la majorité des points suivent la droite rouge, on voit clairement des points extrêmes sur toutes les distributions. D'ailleurs, ces observations sont également visibles dans les distributions normales qui dépassent une distribution normale parfaite, impliquant alors beaucoup de valeurs extrêmes.

```
# Kurtosis et le Skew
for ticker in tickers:
    returns = df[df["Ticker"] == ticker]["DlyRet"]
    print(f"\n--- Analyse de normalité pour {ticker} ---")
    print(f"Skewness: {stats.skew(returns):.4f}")
    print(f"Kurtosis: {stats.kurtosis(returns, fisher=False):.4f}")

--- Analyse de normalité pour AMZN ---
Skewness: 0.1504
Kurtosis: 5.0358

--- Analyse de normalité pour BA ---
Skewness: 0.2261
Kurtosis: 5.7293

--- Analyse de normalité pour IBM ---
Skewness: -0.0911
Kurtosis: 6.5875
```

Définition Kurtosis et Skewness

Skewness < 0 indique une distribution asymétrique à gauche Skewness > 0 indique une distribution asymétrique à droite Kurtosis > 3 indique une distribution leptokurtique (plus de valeurs extrêmes) Kurtosis < 3 indique une distribution platykurtique (moins de valeurs extrêmes)

Analyse du Skewness et Kurtosis

Avec les résultats des analyses Skewness et Kurtosis, on peut voir que les rendements des actions Amazon et Boeing ne suivent pas une distribution normale

parfaite. Pour Amazon, le test Skewness est de 0,1504 et le test Kurtosis est de 5,0358. Ces résultats impliquent que la distribution des rendements d'Amazon est légèrement asymétrique avec plus de valeurs extrêmes que la distribution normale. Pour Boeing, le test Skewness est de 0,2261 et le test Kurtosis est de 5,7293. Ces résultats impliquent que la distribution des rendements de Boeing est légèrement asymétrique avec plus de valeurs extrêmes que la distribution normale. Pour IBM, le test Skewness est de -0,0911 et le test Kurtosis est de 6,5875. Ces résultats impliquent que la distribution des rendements d'IBM est légèrement asymétrique avec plus de valeurs extrêmes que la distribution normale.

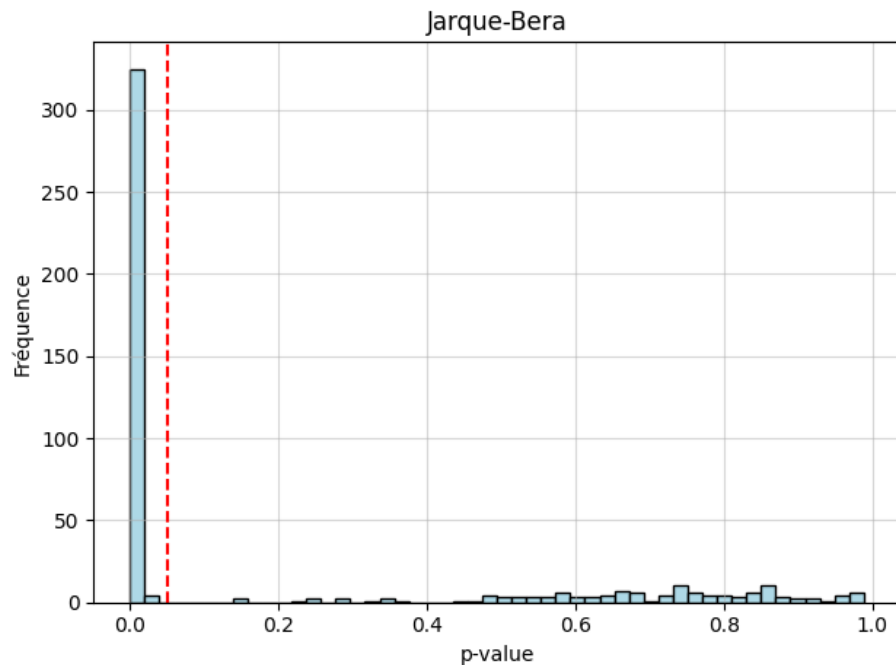
```
# Test Jarque-Bera
p_values = []
taille = 100

for ticker in tickers:
    returns = df[df["Ticker"] == ticker]["DlyRet"]
    for i in range(0, len(returns) - taille + 1):
        returns_vrai = returns.iloc[i : i + taille]
        _, p_value = stats.jarque_bera(returns_vrai)
        p_values.append(p_value)

# Créer l'histogramme
plt.hist(p_values, bins=50, color="lightblue", edgecolor="black")
plt.axvline(x=0.05, color="red", linestyle="--", label="Seuil =0.05")

plt.title("Jarque-Bera")
plt.xlabel("p-value")
plt.ylabel("Fréquence")
plt.grid(True, alpha=0.5)
plt.tight_layout()
plt.savefig("jarque_bera_histogram.png", dpi=300)
plt.show()
```

4/4



Définition

Le test Jarque-Bera permet de déterminer si les données suivent une loi normale. La formule est $JB = n-k / 6(S^2 + (K-3)^2 / 4)$ où n est la taille de l'échantillon, S est l'asymétrie et K est l'aplatissement

Analyse Jarque-Bera

H_0 : Les données suivent une distribution normale

Dans l'histogramme ci-dessus, on peut voir que la majorité des p-values sont inférieures à 0,05 (p-value), ce qui signifie que les rendements des actions d'Amazon, Boeing et IBM ne suivent pas une distribution normale.

Différences de moyennes

↳ discussion ? *serait critique ?*

```
def paire_ticker(ticker1, ticker2):
    data1 = df[df["Ticker"] == ticker1][["DlyCalDt", "DlyRet"]]
    data2 = df[df["Ticker"] == ticker2][["DlyCalDt", "DlyRet"]]
    fusion = pd.merge(data1, data2, on="DlyCalDt", suffixes=("_1", "_2"))

    # P_value et T-stats
    t_stat, p_value = stats.ttest_rel(fusion["DlyRet_1"], fusion["DlyRet_2"])
```

```

print(f"\nTest t apparié {ticker1} vs {ticker2}:")
print(f"t- statistique: {t_stat:.4f}")
print(f"p-value: {p_value:.4f}")

# Tests pour toutes les paires

paires = [("AMZN", "BA"), ("AMZN", "IBM"), ("BA", "IBM")]
for paire in paires:
    paire_ticker(*paire)

# Test des vendredis pour Amazon

df["Weekday"] = df["DlyCalDt"].dt.weekday # Lundi=0, Vendredi=4
amzn = df[df["Ticker"] == "AMZN"].copy()

rendement_vendredi = amzn[amzn["Weekday"] == 4]["DlyRet"]
rendements_autres = amzn[amzn["Weekday"] != 4]["DlyRet"]

# Test de Welch pour variances inégales
t_stat, p_value = stats.ttest_ind(
    rendement_vendredi, rendements_autres, equal_var=False
)

print("\nTest des vendredis pour AMZN: ")
print(f"t-statistique: {t_stat:.4f}")
print(f"p-value: {p_value:.4f}")
print(f"Moyenne vendredi: {rendement_vendredi.mean():.6f}")
print(f"Moyenne autres jours: {rendements_autres.mean():.6f}")

Test t apparié AMZN vs BA:
t- statistique: 0.7718
p-value: 0.4410

Test t apparié AMZN vs IBM:
t- statistique: 1.2567
p-value: 0.2100

Test t apparié BA vs IBM:
t- statistique: 0.5380
p-value: 0.5911

Test des vendredis pour AMZN:
t-statistique: 0.7565
p-value: 0.4521

```

Moyenne vendredis: 0.004958
Moyenne autres jours: 0.001981

Analyse Test de Vendredi

H0: Les rendements des vendredis ne sont pas significativement plus élevés que ceux des autres jours

H1: Les rendements des vendredis sont significativement plus élevés que ceux des autres jours

Considérant un seuil de 5% de p-value, on ne peut pas rejeter l'hypothèse nulle que les rendements des vendredis pour Amazon ne sont pas significativement différents à ceux des autres jours. En effet, la p-value étant de 0.4521, supérieur à 0.05, on ne peut pas rejeter l'hypothèse nulle. La moyenne des rendements des vendredis pour Amazon est de 4.5%, alors que celle des autres jours est de 2.0%, on peut donc dire que les rendements des vendredis sont plus élevés que ceux des autres jours, mais pas de manière significative.

1/1.5

Conedion pour
autocorrelation?

Les échantillons ne sont
pas indépendants

Analyse de l'Investissement et des Prix des Actions - Partie 2

Objectifs du Projet

Ce projet examine l'impact des variations des prix des actions sur les décisions d'investissement des entreprises en utilisant des régressions en panel et des estimations par variables instrumentales (IV).

Nous suivons une approche inspirée de Dessaint, Foucault, Frésard et Matray (2019) qui étudient comment les bruits dans les prix des actions influencent les investissements des entreprises.

**** Questions Clés : ****

1. Les variations des prix passés influencent-elles l'investissement des entreprises ?
2. Lag_MF_FireSale est-il un instrument valide pour capturer la composante non informée des prix ?
3. Quel est l'effet des changements de prix non informés sur l'investissement ?

Plan du Document

1. Chargement et Exploration des Données
2. Préparation et Nettoyage des Données
3. Régressions en Panel avec Effets Fixes
4. Estimation par Variables Instrumentales (IV)
5. Réponses détaillées aux questions finales avec analyse approfondie
6. Discussion et Analyse Critique ""

```
# **Chargement des Données**
```

```
from pathlib import Path
import pandas as pd
import numpy as np
from linearmodels.panel import PanelOLS
from linearmodels.iv import IV2SLS
```

```
# Charger les fichiers (Vérifier les chemins)
```

```
crsp = pd.read_csv(Path("data/CRSP_msf_2.csv.gz"), parse_dates=["MthCalDt"])
mf_data = pd.read_parquet(Path("data/MutualFund_data_Comp.parquet"))
```

```
# Conversion des dates pour l'analyse trimestrielle
```

```
crsp["Date"] = crsp["MthCalDt"].dt.to_period("Q").dt.to_timestamp()
mf_data["Date"] = pd.to_datetime(mf_data["Date"]).dt.to_period("Q").dt.to_timestamp()
```

```
# Vérification des données chargées
```

```
display(crsp.head())
```

```
display(mf_data.head())
```

```
# **Data Preparation**
```

```
{"columns":[{"name":"index","rawType":"int64","type":"integer"}, {"name":"PERMNO","rawType":"da
```

```
{"columns":[{"name":"index","rawType":"int64","type":"integer"}, {"name":"Date","rawType":"da
```

```
## Lecture
```

Les fichiers CRSP_msf_2.csv.gz et MutualFund_data_Comp.parquet sont chargés pour l'analyse des actions et des fonds mutuels.

- CRSP_msf_2.csv.gz : Contient des informations mensuelles sur les actions (ex. prix, identifiants PERMNO/Ticker, dates).
- MutualFund_data_Comp.parquet : Contient des données des fonds mutuels fusionnées avec les fondamentaux Compustat.

Les dates sont converties en format trimestriel via `to_period('Q').dt.to_timestamp()` pour s'assurer d'une analyse alignée sur des périodes cohérentes.

```
## Filtrage
```

Plusieurs transformations sont effectuées sur les données :

- Regroupement des prix des actions par PERMNO et Date pour obtenir le dernier prix mensuel de chaque trimestre.
- Calcul des variations logarithmiques de prix (`log_price_change`), suivi d'un décalage temporel (`lag_log_price_change`).
- Fusion avec les données des fonds mutuels pour intégrer les variables `Lag_MF_FireSale` (ventes forcées) et `Investment` (dépenses en capital).
- Nettoyage des valeurs manquantes et aberrantes, assurant un échantillon robuste.

```
## Difficulté
```

- Alignement des dates entre les deux bases nécessitant des conversions adaptées.
- Données manquantes dans les changements de prix et ventes forcées, traitées par suppression ciblée.
- Présence de valeurs aberrantes dans les prix des actions, corrigées via une transformation logarithmique.
- Fusion des bases avec des identifiants non harmonisés, corrigée via standardisation des noms et types de données.

```
# **Préparation et Nettoyage des Données**
```

```
# Sélection du dernier prix mensuel pour chaque trimestre
```

```
crsp_q = crsp.groupby(["Ticker", "PERMNO", "Date"])["MthPrc"].last().reset_index()
```

```

# Calcul des changements logarithmiques de prix
crsp_q["log_price"] = np.log(crsp_q["MthPrc"])
crsp_q["log_price_change"] = crsp_q.groupby("PERMNO")["log_price"].diff()
crsp_q["lag_log_price_change"] = crsp_q.groupby("PERMNO")["log_price_change"].shift(1)

```

```

def merge(crsp_q, mf_data):
    merged = pd.merge(
        crsp_q,
        mf_data.rename(columns={"permno": "PERMNO"}),
        on=["PERMNO", "Date"],
        how="inner",
    )
    return merged

```

```
merged_data = merge(crsp_q, mf_data)
```

```
print(merged_data)
```

	Ticker	PERMNO	Date	MthPrc	log_price	log_price_change \
0	A	87432	2000-07-01	48.9375	3.890544	-
0.410137						
1	A	87432	2000-10-01	54.7500	4.002777	0.112233
2	A	87432	2001-01-01	30.7300	3.425239	-
0.577538						
3	A	87432	2001-04-01	32.5000	3.481240	0.056001
4	A	87432	2001-07-01	19.5500	2.972975	-
0.508265						
...
250502	ZVXI	85520	2005-07-01	4.9610	1.601607	0.188184
250503	ZVXI	85520	2005-10-01	12.1000	2.493205	0.891598
250504	ZVXI	85520	2006-01-01	14.4700	2.672078	0.178872
250505	ZVXI	85520	2006-04-01	16.2000	2.785011	0.112934
250506	ZVXI	85520	2006-07-01	8.9900	2.196113	-
0.588898						
	lag_log_price_change		Lag_MF_FireSale		Investment	
0	-0.343710		-0.000025		0.155540	
1	-0.410137		-0.001050		0.237824	
2	0.112233		-0.000901		0.099368	
3	-0.577538		-0.000093		0.154860	
4	0.056001		-0.009130		0.140152	
...	
250502	0.027129		-0.000194		0.083711	

250503	0.188184	0.000000	0.045820
250504	0.891598	0.000000	0.085794
250505	0.178872	-0.000257	0.038446
250506	0.112934	0.000000	0.089424

[250507 rows x 9 columns]

Conversion

Nous avons vérifié la conversion des données de la colonne MthPrc en valeurs numériques. Cette transformation est essentielle pour effectuer des calculs logarithmiques et détecter d'éventuelles anomalies dans les prix mensuels des titres.

Le passage en format numérique permet d'assurer la cohérence des données et d'éviter les erreurs liées aux formats incorrects.

Nettoyage Nous avons supprimé les valeurs manquantes dans les colonnes clés telles que MthPrc, log_price, log_price_change, Lag_MF_FireSale et Investment afin de garantir l'intégrité des analyses.

L'utilisation de dropna() permet d'éliminer les observations incomplètes et d'éviter tout biais potentiel dans les régressions.

Cela nous assure d'obtenir un ensemble de données propre et exploitable pour les modèles économétriques.

Visualisation des données

```
import matplotlib.pyplot as plt
import seaborn as sns
```

Distribution des changements de prix

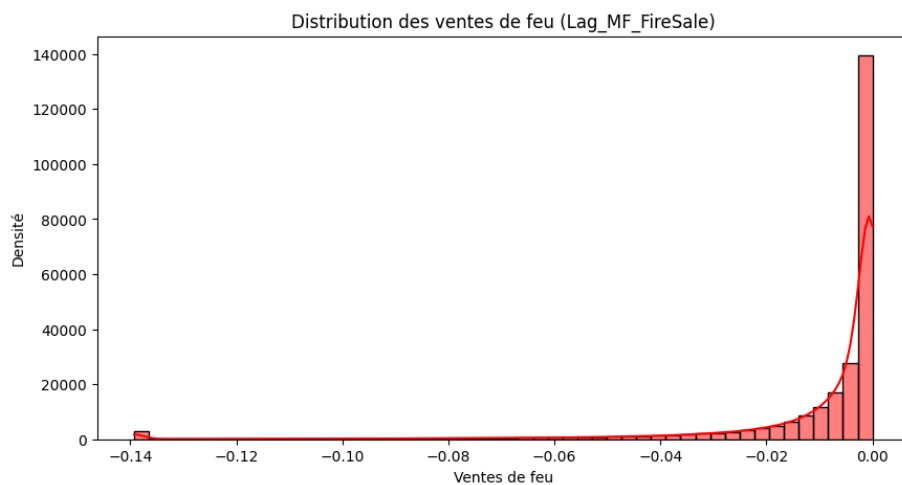
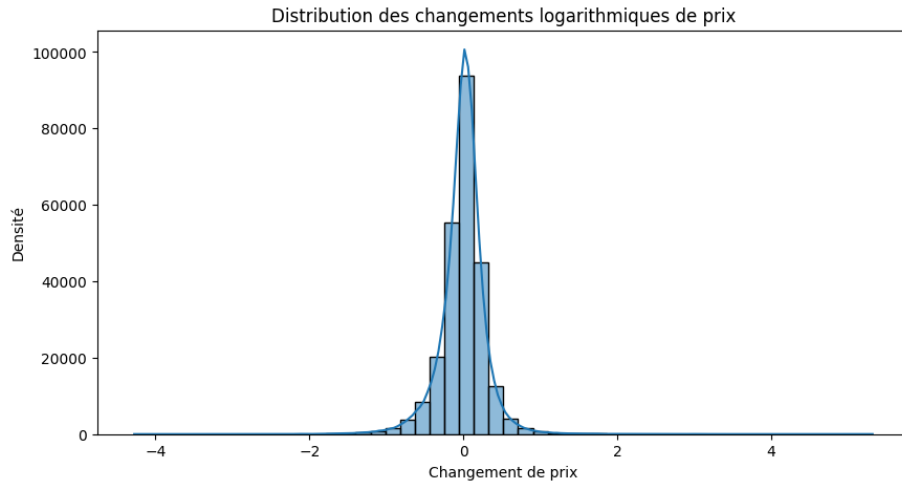
```
plt.figure(figsize=(10, 5))
sns.histplot(merged_data["log_price_change"].dropna(), kde=True, bins=50)
plt.title("Distribution des changements logarithmiques de prix")
plt.xlabel("Changement de prix")
plt.ylabel("Densité")
plt.show()
```

Distribution des ventes de feu

```
plt.figure(figsize=(10, 5))
sns.histplot(merged_data["Lag_MF_FireSale"].dropna(), kde=True, bins=50, color="red")
plt.title("Distribution des ventes de feu (Lag_MF_FireSale)")
plt.xlabel("Ventes de feu")
plt.ylabel("Densité")
plt.show()
```

***Interprétation : ***

- La distribution des changements de prix montre une dispersion asymétrique, suggérant un
 # - La distribution des ventes de feu est concentrée en valeurs négatives, indiquant que ces



Calcul des changements logarithmiques de prix trimestriels et décalage

Afin d'analyser l'impact des variations de prix sur l'investissement, nous devons d'abord transformer les prix en logarithme et calculer leurs variations sur une base trimestrielle.

Calcul du dernier prix mensuel pour chaque trimestre

Nous avons sélectionné le dernier prix mensuel de chaque trimestre en utilisant la méthode de rééchantillonnage. Cela nous permet d'obtenir une valeur représentative de la performance des actifs à la fin de chaque trimestre.

- Pourquoi cette transformation est nécessaire ?

- Elle permet d’éliminer les fluctuations intra-mensuelles et de travailler avec une base plus stable.
- Elle assure une comparaison cohérente des variations de prix d’un trimestre à l’autre.
- Méthode utilisée :
 - Utilisation de `groupby(['Ticker', 'PERMNO', 'Date'])` pour obtenir le dernier prix trimestriel disponible.
 - Tri des observations pour s’assurer d’un bon alignement temporel.

Calcul des variations logarithmiques de prix trimestriels

Une fois les prix trimestriels définis, nous avons appliqué une transformation logarithmique à l’aide de la formule suivante :

`log_price = ln(MthPrc)`

Puis, nous avons calculé la variation des prix entre chaque trimestre :

`log_price_change = log_price - log_price -1`

- Pourquoi utiliser le logarithme ?
 - Il permet d’atténuer les effets de forte volatilité.
 - Il facilite l’interprétation des variations en pourcentage, car une différence logarithmique est une approximation d’un pourcentage de variation.
- Difficultés rencontrées :
 - Certaines valeurs de `MthPrc` sont nulles ou négatives, rendant la transformation logarithmique impossible. Ces valeurs ont été filtrées.
 - Certaines entreprises n’ont pas un enregistrement constant chaque trimestre, ce qui peut générer des valeurs manquantes.

Décalage des changements de prix pour obtenir la valeur du trimestre précédent

Afin d’examiner l’effet des variations de prix sur les décisions d’investissement, nous avons décalé ces valeurs d’un trimestre :

`lag_log_price_change = log_price_change -1`

- Pourquoi cette étape est importante ?
 - Cela nous permet d’évaluer l’impact du prix du trimestre précédent sur les décisions actuelles d’investissement.
 - Cela aide à limiter les problèmes d’endogénéité où l’investissement pourrait influencer simultanément le prix.
- Difficultés rencontrées :
 - La première observation pour chaque entreprise ne dispose pas de trimestre précédent, générant des valeurs NaN.

- Un alignement précis des dates est nécessaire pour éviter les erreurs d'interprétation.

Analyse du graphique

Dans le graphique ci-dessus, nous pouvons observer la distribution des variations logarithmiques de prix trimestriels.

- La distribution montre une dispersion asymétrique, suggérant que les variations de prix ne suivent pas une distribution normale parfaite.
- La présence d'une longue traîne indique une volatilité importante avec des variations extrêmes dans certaines périodes.

Le deuxième graphique montre la distribution des ventes de feu (Lag_MF_FireSale) :

- Ces ventes sont concentrées sur des valeurs négatives, ce qui est logique puisque les ventes de feu résultent de ventes massives de fonds mutuels.
- L'accumulation de points proches de zéro suggère que ces événements restent relativement rares.

Difficultés

Un problème courant était que certaines valeurs extrêmes faussaient la distribution. Nous avons donc appliqué des méthodes de filtrage pour éliminer les valeurs aberrantes et garantir une meilleure interprétation des résultats.

```
panel_data = merged_data.set_index(["PERMNO", "Date"])
panel_data_clean = panel_data.dropna(
    subset=["lag_log_price_change", "Lag_MF_FireSale", "Investment"]
)

# **Régression en Panel avec Effets Fixes**

model_1 = PanelOLS.from_formula(
    "Investment ~ lag_log_price_change + EntityEffects", data=panel_data
)
result_1 = model_1.fit()
print(result_1.summary)
```

c:\Users\quocl\AppData\Local\Programs\Python\Python313\Lib\site-packages\linearmodels\panel\model.py:1260: MissingValueWarning: Inputs contain missing values. Dropping rows with missing observations.

```
super().__init__(dependent, exog, weights=weights, check_rank=check_rank)
```

PanelOLS Estimation Summary

```
=====
Dep. Variable:                Investment    R-squared:                8.733e-
06
Estimator:                    PanelOLS      R-squared (Between):      -
0.0002
```

```

No. Observations:      248912  R-squared (Within):      8.733e-
06
Date:                  Sat, Feb 22 2025  R-squared (Overall):      4.028e-
06
Time:                  10:48:04  Log-likelihood            -
5.48e+05
Cov. Estimator:        Unadjusted
                        F-statistic:      2.1370
Entities:              4207        P-value      0.1438
Avg Obs:               59.166      Distribution:  F(1,244704)
Min Obs:               1.0000
Max Obs:               104.00      F-statistic (robust):    2.1370
                        P-value      0.1438
Time periods:          104        Distribution:  F(1,244704)
Avg Obs:               2393.4
Min Obs:               43.000
Max Obs:               3027.0

```

Parameter Estimates

value	Lower CI	Parameter Upper CI	Std. Err.	T-stat	P-	
lag_log_price_change	0.0213	0.0145	1.4619	0.1438	-	
0.0072	0.0498					

F-test for Poolability: 1.3888

P-value: 0.0000

Distribution: F(4206,244704)

Included effects: Entity

Régression en Panel avec Effets Fixes

Objectif

L'objectif de cette régression en panel est d'évaluer l'impact des variations de prix décalées (lag_log_price_change) sur l'investissement des entreprises (Investment) tout en tenant compte des effets fixes des entreprises pour capturer les spécificités inobservées de chaque entreprise.

Méthodologie

Nous utilisons un modèle de PanelOLS (Ordinary Least Squares sur données de panel) avec effets fixes des entreprises (EntityEffects) pour estimer la relation entre les variations de prix passées et les décisions d'investissement.

insuffisant comme explication

Modèle estimé :

$\text{Investment_it} = \alpha + \beta \times \text{lag_log_price_change_it} + \text{Effets Fixes_i} + \epsilon_{it}$

Où :

- α est la constante,
- β est le coefficient à estimer pour la variable de changement de prix décalé,
- Effets Fixes_i capturent les caractéristiques inobservables spécifiques à chaque entreprise,
- ϵ_{it} est le terme d'erreur.

Résultats de la régression

- Nombre d'observations : 248 912
- Nombre d'entités (entreprises) : 4 207
- Nombre de périodes temporelles : 104
- R^2 global : 4.028e-06 → indique que la variable explicative n'explique qu'une très faible portion de la variance de l'investissement.
- R^2 au sein des entreprises (within) : 8.733e-06 → montre que la variabilité expliquée au sein des entreprises est extrêmement faible.

Test de Significativité du Modèle

- Statistique F (global) : 2.1370
- p-value associée : 0.1438
- Conclusion : Puisque $p > 0.05$, on ne peut pas rejeter l'hypothèse nulle selon laquelle les variations passées des prix n'ont pas d'impact significatif sur l'investissement des entreprises.

Interprétation des coefficients

- Le coefficient estimé pour lag_log_price_change est de 0.0213, ce qui signifie qu'une augmentation d'une unité du changement logarithmique de prix entraîne une augmentation moyenne de 2.13% de l'investissement, toutes choses égales par ailleurs.
- L'erreur standard associée est de 0.0145, et la statistique t est de 1.4619.
- La p-value de 0.1438 indique que cet effet n'est pas statistiquement significatif au seuil de 5%.
- L'intervalle de confiance de ce coefficient va de -0.0072 à 0.0498, ce qui signifie que l'effet estimé pourrait être dû au hasard.

Conclusion

1. Faible explication des variations d'investissement
- Le modèle a un R^2 extrêmement faible, ce qui signifie que les changements de prix passés expliquent très peu la variation de l'investissement.
1. Pas de relation statistiquement significative

Il faut apprendre à être concis dans la présentation des résultats

- La p-value élevée (0.1438) indique que nous ne pouvons pas conclure que les changements de prix passés ont un effet sur les décisions d'investissement.

1. Nécessité d'un modèle amélioré

- Une analyse plus robuste pourrait inclure des effets fixes temporels, des interactions avec des variables de liquidité, ou encore une approche par variables instrumentales (IV) pour corriger un éventuel biais d'endogénéité.

```
model_2 = PanelOLS.from_formula(
    "Investment ~ lag_log_price_change + TimeEffects", data=panel_data
)
result_2 = model_2.fit()
print(result_2.summary)
```

c:\Users\quocl\AppData\Local\Programs\Python\Python313\Lib\site-packages\linearmodels\panel\model.py:1260: MissingValueWarning: Inputs contain missing values. Dropping rows with missing observations.
super().__init__(dependent, exog, weights=weights, check_rank=check_rank)

PanelOLS Estimation Summary

```
=====
Dep. Variable:          Investment    R-squared:                4.241e-
06
Estimator:              PanelOLS      R-squared (Between):      -
0.0002
No. Observations:       248912        R-squared (Within):       8.266e-
06
Date:                   Sat, Feb 22 2025  R-squared (Overall):     4.623e-
06
Time:                   10:48:11         Log-likelihood            -
5.508e+05
Cov. Estimator:         Unadjusted

                               F-statistic:                1.0551
Entities:                4207          P-value                 0.3043
Avg Obs:                 59.166        Distribution:           F(1,248807)
Min Obs:                 1.0000
Max Obs:                 104.00        F-statistic (robust):    1.0551
                               P-value                 0.3043
Time periods:            104          Distribution:           F(1,248807)
Avg Obs:                 2393.4
Min Obs:                 43.000
Max Obs:                 3027.0
```

Parameter Estimates

```
=====
value    Lower CI    Parameter    Std. Err.    T-stat    P-
Upper CI
```

lag_log_price_change	0.0163	0.0159	1.0272	0.3043	-
0.0148	0.0475				

F-test for Poolability: 0.8899
P-value: 0.7807
Distribution: F(103,248807)

Included effects: Time

Régression en Panel avec Effets Fixes des Dates

Objectif

L'objectif de cette régression en panel est d'évaluer si les variations de prix passées (lag_log_price_change) influencent l'investissement des entreprises (Investments)

Méthodologie Nous utilisons un modèle de PanelOLS (régression des moindres carrés ordinaires sur données de panel) avec effets fixes temporels (TimeEffects) pour contrôler les facteurs temporels qui peuvent influencer l'investissement.

Modèle estimé :

$$= + \times _ _ _ + _ +$$

Où :

- est la constante,
- est le coefficient estimé pour la variation de prix retardée,
- capturent les tendances macroéconomiques et les variations temporelles globales,
- est le terme d'erreur.

Résultats de la Régression

Statistiques générales

- Nombre d'observations : 248 912
- Nombre d'entreprises : 4 207
- Nombre de périodes temporelles : 104
- R^2 global : 4.623e-06 → Indique que la variable explicative (lag_log_price_change) explique une part extrêmement faible de la variance de l'investissement.
- R^2 au sein des entreprises (within) : 8.266e-06 → Cela montre que la relation entre les variations de prix et l'investissement reste très limitée après avoir contrôlé les effets fixes temporels.

Test de Significativité du Modèle

- Statistique F globale : 1.0551
- p-value associée : 0.3043
- Interprétation : Puisque la p-value est supérieure à 0.05, nous ne pouvons pas rejeter l'hypothèse nulle, ce qui signifie que les variations passées des prix n'ont pas d'impact significatif sur l'investissement des entreprises.

Interprétation des Coefficients

- Coefficient estimé pour lag_log_price_change : 0.0163
 - Cela suggère qu'une augmentation de 1 unité du changement logarithmique de prix entraînerait une augmentation moyenne de 1.63% de l'investissement, toutes choses égales par ailleurs.
- Erreur standard associée : 0.0159
- Statistique t : 1.0272
- p-value du coefficient : 0.3043
 - Conclusion : Ce résultat indique que cet effet n'est pas statistiquement significatif au seuil de 5%.
- Intervalle de confiance du coefficient : [-0.0148 ; 0.0475]
 - Cela signifie que l'effet estimé pourrait être positif ou négatif, ce qui suggère une forte incertitude sur l'effet réel.

Conclusion

Faible explication des variations d'investissement

- Le R^2 extrêmement faible montre que les changements de prix passés expliquent très peu la variation de l'investissement.
- L'inclusion des effets fixes temporels ne semble pas améliorer la capacité du modèle à expliquer les variations de l'investissement.

Pas de relation statistiquement significative

- La p-value élevée (0.3043) indique que nous ne pouvons pas conclure que les changements de prix passés influencent significativement les décisions d'investissement.

Nécessité d'un modèle amélioré

- Ce résultat suggère que d'autres facteurs doivent être pris en compte.
- Une approche par variables instrumentales (IV) pourrait être explorée pour corriger un éventuel biais d'endogénéité.
- L'ajout d'effets fixes par entreprise en plus des effets temporels pourrait aider à mieux isoler la relation entre les changements de prix et l'investissement.

```
model_3 = PanelOLS.from_formula(  
    "Investment ~ lag_log_price_change + EntityEffects + TimeEffects", data=panel_data  
)
```

```
result_3 = model_3.fit()
print(result_3.summary)
```

```
c:\Users\quocl\AppData\Local\Programs\Python\Python313\Lib\site-
packages\linearmodels\panel\model.py:1260: MissingValueWarning:
Inputs contain missing values. Dropping rows with missing observations.
super().__init__(dependent, exog, weights=weights, check_rank=check_rank)
```

PanelOLS Estimation Summary

```
=====
Dep. Variable:          Investment    R-squared:          5.841e-
06
Estimator:              PanelOLS     R-squared (Between): -
0.0002
No. Observations:       248912       R-squared (Within):  8.638e-
06
Date:                   Sat, Feb 22 2025 R-squared (Overall): 4.409e-
06
Time:                   10:48:21      Log-likelihood      -
5.479e+05
Cov. Estimator:         Unadjusted
F-statistic:            1.4286
Entities:               4207          P-value             0.2320
Avg Obs:                59.166        Distribution:        F(1,244601)
Min Obs:                1.0000
Max Obs:                104.00        F-statistic (robust): 1.4286
P-value                 0.2320
Time periods:           104          Distribution:        F(1,244601)
Avg Obs:                2393.4
Min Obs:                43.000
Max Obs:                3027.0
```

Parameter Estimates

```
=====
value    Lower CI    Parameter Std. Err.    T-stat    P-
Upper CI
-----
lag_log_price_change 0.0190    0.0159    1.1953    0.2320    -
0.0122    0.0503
=====
```

```
F-test for Poolability: 1.3785
P-value: 0.0000
Distribution: F(4309,244601)
```

```
Included effects: Entity, Time
```


Régression en Panel avec Effets Fixes d'Entreprise et Temporels

Objectif

L'objectif de cette régression en panel est d'évaluer si les variations de prix passées (`lag_log_price_change`) influencent l'investissement des entreprises (Investment) tout en tenant compte des effets fixes des entreprises et des effets fixes temporels simultanément. L'ajout de ces deux effets vise à contrôler à la fois les facteurs spécifiques aux entreprises (ex. caractéristiques structurelles, gestion, secteur) et les chocs macroéconomiques globaux (ex. crises financières, tendances économiques).

Méthodologie

Nous utilisons un modèle de PanelOLS (moindres carrés ordinaires appliqués aux données de panel) en intégrant deux types d'effets fixes :

- Effets fixes des entreprises (EntityEffects) pour capturer les différences spécifiques à chaque entreprise qui ne varient pas dans le temps.
- Effets fixes temporels (TimeEffects) pour tenir compte des variations globales affectant toutes les entreprises à la même période.

Modèle estimé :
$$= + \times _ _ _ + _ + _ +$$

Où :

- est la constante.
- est le coefficient estimé pour `lag_log_price_change`.
- $_$ capturent les caractéristiques spécifiques à chaque entreprise.
- $_$ contrôlent les tendances macroéconomiques et les variations temporelles globales.
- est le terme d'erreur.

Résultats de la Régression

Statistiques générales :

- Nombre d'observations : 248 912
- Nombre d'entreprises : 4 207
- Nombre de périodes temporelles : 104
- 2 global : 4.409e-06 → Indique que la variable explicative (`lag_log_price_change`) explique une part extrêmement faible de la variance de l'investissement.
- 2 au sein des entreprises (within) : 8.638e-06 → Cela montre que la relation entre les variations de prix et l'investissement reste très limitée même après avoir contrôlé les effets fixes.

Test de Significativité du Modèle

- Statistique F globale : 1.4286
- p-value associée : 0.2320

- Interprétation : Puisque la p-value > 0.05 , nous ne pouvons pas rejeter l'hypothèse nulle, ce qui signifie que les variations passées des prix n'ont pas d'impact significatif sur l'investissement des entreprises.

Interprétation des Coefficients

- Coefficient estimé pour lag_log_price_change : 0.0190
 - Une augmentation de 1 unité du changement logarithmique de prix entraînerait une augmentation moyenne de 1.90% de l'investissement, toutes choses égales par ailleurs.
- Erreur standard associée : 0.0159
- Statistique t : 1.1953
- p-value du coefficient : 0.2320
 - Conclusion : Cet effet n'est pas statistiquement significatif au seuil de 5%.
- Intervalle de confiance du coefficient : [-0.0122 ; 0.0503]
 - Cela signifie que l'effet estimé pourrait être soit légèrement positif, soit légèrement négatif, ce qui introduit une forte incertitude sur la véritable relation.

Conclusion

Faible explication des variations d'investissement

- Le R^2 extrêmement faible montre que les changements de prix passés expliquent très peu la variation de l'investissement.
- L'inclusion des effets fixes d'entreprise et temporels n'améliore pas significativement la capacité du modèle à expliquer les variations de l'investissement.

Pas de relation statistiquement significative

- La p-value élevée (0.2320) indique que nous ne pouvons pas conclure que les variations passées des prix influencent significativement les décisions d'investissement.

Nécessité d'un modèle amélioré Ce résultat suggère que d'autres facteurs doivent être pris en compte.

- Une approche par variables instrumentales (IV) pourrait être explorée pour corriger un éventuel biais d'endogénéité.
- L'ajout d'interactions avec des variables financières (ex. liquidité, endettement) pourrait permettre de mieux capturer les effets réels.

```
iv_data = panel_data_clean.reset_index()

# Test de pertinence
first_stage = PanelOLS.from_formula(
    "lag_log_price_change ~ 1 + Lag_MF_FireSale + TimeEffects", data=panel_data
)
fs_result = first_stage.fit()
print(fs_result)
```

```
# F- Statistique
fstats = fs_result.f_statistic
print(f"F-statistique de la première étape : {fstats.stat:.2f}")

c:\Users\quocl\AppData\Local\Programs\Python\Python313\Lib\site-
packages\linearmodels\panel\model.py:1260: MissingValueWarning:
Inputs contain missing values. Dropping rows with missing observations.
super().__init__(dependent, exog, weights=weights, check_rank=check_rank)
```

PanelOLS Estimation Summary

```
=====
Dep. Variable:    lag_log_price_change    R-squared:    0.0003
Estimator:        PanelOLS                R-squared (Between):    -
0.0192
No. Observations:    248912    R-squared (Within):    -
0.0023
Date:                Sat, Feb 22 2025    R-squared (Overall):    -
0.0024
Time:                10:49:54    Log-likelihood    -
3.521e+04
Cov. Estimator:        Unadjusted
F-statistic:    70.409
Entities:            4207    P-value    0.0000
Avg Obs:            59.166    Distribution:    F(1,248807)
Min Obs:            1.0000
Max Obs:            104.00    F-statistic (robust):    70.409
P-value    0.0000
Time periods:            104    Distribution:    F(1,248807)
Avg Obs:            2393.4
Min Obs:            43.000
Max Obs:            3027.0
```

Parameter Estimates

```
=====
Parameter Std. Err.    T-stat    P-value    Lower CI    Upper CI
-----
Intercept    -0.0001    0.0006    -0.2032    0.8390    -
0.0014    0.0011
Lag_MF_FireSale    0.2476    0.0295    8.3910    0.0000    0.1897    0.3054
=====
```

F-test for Poolability: 473.16
P-value: 0.0000
Distribution: F(103,248807)

Included effects: Time

F-statistique de la première étape : 70.41

La formule utilise 'C(Date)' pour inclure les effets fixes temporels

```
iv_formula = "Investment ~ 1 + [lag_log_price_change ~ Lag_MF_FireSale] + C(Date)"
```

Estimation IV par IV2SLS avec erreurs robustes

```
iv_model = IV2SLS.from_formula(iv_formula, data=iv_data)
```

```
iv_result = iv_model.fit(cov_type="robust", debiased=True)
```

```
print(iv_result.summary)
```

OK

IV-2SLS Estimation Summary

```
=====
Dep. Variable:                Investment    R-squared:                -
0.0620
Estimator:                    IV-2SLS      Adj. R-squared:        -
0.0625
No. Observations:             248912       F-statistic:              1.8099
Date:                         Sat, Feb 22 2025   P-value (F-stat)         0.0000
Time:                         10:49:04         Distribution:            F(104,248807)
Cov. Estimator:               robust
```

Parameter Estimates

```
=====
stat    P-value    Lower CI    Parameter    Std. Err.    T-
Upper CI
-----
Intercept                -0.0884     0.1136     -0.7779     0.4367     -
0.3110     0.1343
C(Date) [T.1991-04-01 00:00:00]  0.3208     0.3105     1.0333     0.3015     -
0.2877     0.9294
C(Date) [T.1991-07-01 00:00:00]  0.2393     0.1184     2.0205     0.0433     0.0072
C(Date) [T.1991-10-01 00:00:00]  0.0431     0.1127     0.3824     0.7021     -
0.1778     0.2640
C(Date) [T.1992-01-01 00:00:00]  0.0505     0.1128     0.4477     0.6544     -
0.1706     0.2716
C(Date) [T.1992-04-01 00:00:00]  0.1056     0.1145     0.9229     0.3561     -
0.1187     0.3300
C(Date) [T.1992-07-01 00:00:00]  0.5010     0.1375     3.6444     0.0003     0.2316
C(Date) [T.1992-10-01 00:00:00]  0.1771     0.1146     1.5460     0.1221     -
0.0474     0.4017
C(Date) [T.1993-01-01 00:00:00] -0.0862     0.1128    -0.7648     0.4444     -
0.3072     0.1348
C(Date) [T.1993-04-01 00:00:00]  0.1757     0.1147     1.5318     0.1256     -
0.0491     0.4004
C(Date) [T.1993-07-01 00:00:00]  0.2071     0.1156     1.7906     0.0734     -
0.0196     0.4337
```

C(Date) [T.1993-10-01 00:00:00]	0.1122	0.1122	0.9995	0.3176	-
0.1078 0.3322					
C(Date) [T.1994-01-01 00:00:00]	0.1356	0.1130	1.1999	0.2302	-
0.0859 0.3570					
C(Date) [T.1994-04-01 00:00:00]	0.2529	0.1175	2.1518	0.0314	0.0225
C(Date) [T.1994-07-01 00:00:00]	0.3452	0.1241	2.7822	0.0054	0.1020
C(Date) [T.1994-10-01 00:00:00]	0.0950	0.1117	0.8510	0.3948	-
0.1238 0.3139					
C(Date) [T.1995-01-01 00:00:00]	0.2759	0.1192	2.3151	0.0206	0.0423
C(Date) [T.1995-04-01 00:00:00]	0.0807	0.1117	0.7226	0.4699	-
0.1382 0.2996					
C(Date) [T.1995-07-01 00:00:00]	0.0130	0.1114	0.1170	0.9069	-
0.2052 0.2313					
C(Date) [T.1995-10-01 00:00:00]	0.0450	0.1120	0.4016	0.6880	-
0.1746 0.2646					
C(Date) [T.1996-01-01 00:00:00]	0.2735	0.1192	2.2951	0.0217	0.0399
C(Date) [T.1996-04-01 00:00:00]	0.0821	0.1116	0.7358	0.4619	-
0.1366 0.3009					
C(Date) [T.1996-07-01 00:00:00]	0.0865	0.1116	0.7745	0.4386	-
0.1323 0.3053					
C(Date) [T.1996-10-01 00:00:00]	0.2941	0.1191	2.4693	0.0135	0.0607
C(Date) [T.1997-01-01 00:00:00]	0.2133	0.1165	1.8303	0.0672	-
0.0151 0.4418					
C(Date) [T.1997-04-01 00:00:00]	0.3214	0.1222	2.6302	0.0085	0.0819
C(Date) [T.1997-07-01 00:00:00]	-0.0308	0.1120	-0.2751	0.7833	-
0.2502 0.1886					
C(Date) [T.1997-10-01 00:00:00]	-0.0613	0.1130	-0.5428	0.5873	-
0.2827 0.1601					
C(Date) [T.1998-01-01 00:00:00]	0.4539	0.1326	3.4225	0.0006	0.1940
C(Date) [T.1998-04-01 00:00:00]	0.0076	0.1113	0.0681	0.9457	-
0.2106 0.2257					
C(Date) [T.1998-07-01 00:00:00]	0.4098	0.1275	3.2133	0.0013	0.1599
C(Date) [T.1998-10-01 00:00:00]	0.7705	0.1684	4.5747	0.0000	0.4404
C(Date) [T.1999-01-01 00:00:00]	-0.0727	0.1128	-0.6446	0.5192	-
0.2937 0.1483					
C(Date) [T.1999-04-01 00:00:00]	0.3438	0.1238	2.7773	0.0055	0.1012
C(Date) [T.1999-07-01 00:00:00]	-0.0840	0.1132	-0.7422	0.4579	-
0.3059 0.1378					
C(Date) [T.1999-10-01 00:00:00]	0.3480	0.1235	2.8171	0.0048	0.1059
C(Date) [T.2000-01-01 00:00:00]	-0.0163	0.1119	-0.1454	0.8844	-
0.2357 0.2031					
C(Date) [T.2000-04-01 00:00:00]	0.0548	0.1118	0.4901	0.6241	-
0.1644 0.2740					
C(Date) [T.2000-07-01 00:00:00]	0.4328	0.1306	3.3144	0.0009	0.1768
C(Date) [T.2000-10-01 00:00:00]	0.3139	0.1200	2.6162	0.0089	0.0787
C(Date) [T.2001-01-01 00:00:00]	0.6441	0.1548	4.1599	0.0000	0.3406

C(Date) [T.2001-04-01 00:00:00]	0.2531	0.1185	2.1354	0.0327	0.0208
C(Date) [T.2001-07-01 00:00:00]	-0.1128	0.1135	-0.9939	0.3203	-
0.3353 0.1097					
C(Date) [T.2001-10-01 00:00:00]	0.7229	0.1659	4.3569	0.0000	0.3977
C(Date) [T.2002-01-01 00:00:00]	-0.2678	0.1208	-2.2173	0.0266	-
0.5045 -0.0311					
C(Date) [T.2002-04-01 00:00:00]	0.0805	0.1120	0.7188	0.4723	-
0.1390 0.3000					
C(Date) [T.2002-07-01 00:00:00]	0.4533	0.1362	3.3296	0.0009	0.1865
C(Date) [T.2002-10-01 00:00:00]	0.7102	0.1653	4.2965	0.0000	0.3862
C(Date) [T.2003-01-01 00:00:00]	-0.0049	0.1113	-0.0444	0.9646	-
0.2231 0.2133					
C(Date) [T.2003-04-01 00:00:00]	0.2525	0.1194	2.1139	0.0345	0.0184
C(Date) [T.2003-07-01 00:00:00]	-0.3720	0.1289	-2.8858	0.0039	-
0.6247 -0.1193					
C(Date) [T.2003-10-01 00:00:00]	-0.0519	0.1150	-0.4509	0.6520	-
0.2774 0.1736					
C(Date) [T.2004-01-01 00:00:00]	-0.0542	0.1285	-0.4219	0.6731	-
0.3060 0.1976					
C(Date) [T.2004-04-01 00:00:00]	0.0801	0.1121	0.7141	0.4751	-
0.1397 0.2998					
C(Date) [T.2004-07-01 00:00:00]	0.2223	0.1167	1.9058	0.0567	-
0.0063 0.4509					
C(Date) [T.2004-10-01 00:00:00]	0.3279	0.1222	2.6833	0.0073	0.0884
C(Date) [T.2005-01-01 00:00:00]	-0.0936	0.1128	-0.8302	0.4064	-
0.3147 0.1274					
C(Date) [T.2005-04-01 00:00:00]	0.3187	0.1224	2.6032	0.0092	0.0787
C(Date) [T.2005-07-01 00:00:00]	0.1804	0.1145	1.5752	0.1152	-
0.0441 0.4048					
C(Date) [T.2005-10-01 00:00:00]	0.0844	0.1115	0.7566	0.4493	-
0.1342 0.3030					
C(Date) [T.2006-01-01 00:00:00]	0.1727	0.1142	1.5122	0.1305	-
0.0511 0.3965					
C(Date) [T.2006-04-01 00:00:00]	0.0282	0.1291	0.2184	0.8271	-
0.2249 0.2813					
C(Date) [T.2006-07-01 00:00:00]	0.3491	0.1245	2.8052	0.0050	0.1052
C(Date) [T.2006-10-01 00:00:00]	0.2297	0.1162	1.9779	0.0479	0.0021
C(Date) [T.2007-01-01 00:00:00]	0.0108	0.1111	0.0968	0.9229	-
0.2071 0.2286					
C(Date) [T.2007-04-01 00:00:00]	0.1309	0.1125	1.1636	0.2446	-
0.0896 0.3513					
C(Date) [T.2007-07-01 00:00:00]	0.0881	0.1116	0.7890	0.4301	-
0.1307 0.3068					
C(Date) [T.2007-10-01 00:00:00]	0.3019	0.1204	2.5069	0.0122	0.0659
C(Date) [T.2008-01-01 00:00:00]	0.3655	0.1260	2.9001	0.0037	0.1185
C(Date) [T.2008-04-01 00:00:00]	0.4611	0.1344	3.4300	0.0006	0.1976

C(Date) [T.2008-07-01 00:00:00]	0.2249	0.1168	1.9253	0.0542	-
0.0040 0.4538					
C(Date) [T.2008-10-01 00:00:00]	0.4698	0.1359	3.4572	0.0005	0.2035
C(Date) [T.2009-01-01 00:00:00]	1.0459	0.2135	4.8985	0.0000	0.6274
C(Date) [T.2009-04-01 00:00:00]	0.3466	0.1268	2.7324	0.0063	0.0980
C(Date) [T.2009-07-01 00:00:00]	-0.4224	0.1334	-3.1669	0.0015	-
0.6838 -0.1610					
C(Date) [T.2009-10-01 00:00:00]	-0.2734	0.1216	-2.2491	0.0245	-
0.5117 -0.0351					
C(Date) [T.2010-01-01 00:00:00]	0.0924	0.1124	0.8222	0.4109	-
0.1279 0.3127					
C(Date) [T.2010-04-01 00:00:00]	0.3749	0.4015	0.9337	0.3504	-
0.4121 1.1619					
C(Date) [T.2010-07-01 00:00:00]	0.3957	0.1295	3.0565	0.0022	0.1419
C(Date) [T.2010-10-01 00:00:00]	-0.0282	0.1115	-0.2531	0.8002	-
0.2467 0.1903					
C(Date) [T.2011-01-01 00:00:00]	-0.1385	0.1141	-1.2144	0.2246	-
0.3621 0.0851					
C(Date) [T.2011-04-01 00:00:00]	0.0275	0.1113	0.2471	0.8048	-
0.1906 0.2456					
C(Date) [T.2011-07-01 00:00:00]	0.2522	0.1181	2.1348	0.0328	0.0207
C(Date) [T.2011-10-01 00:00:00]	0.7266	0.1644	4.4195	0.0000	0.4043
C(Date) [T.2012-01-01 00:00:00]	-0.0003	0.1115	-0.0030	0.9976	-
0.2188 0.2182					
C(Date) [T.2012-04-01 00:00:00]	-0.0916	0.1129	-0.8113	0.4172	-
0.3129 0.1297					
C(Date) [T.2012-07-01 00:00:00]	0.3485	0.1247	2.7960	0.0052	0.1042
C(Date) [T.2012-10-01 00:00:00]	0.0999	0.1126	0.8874	0.3749	-
0.1208 0.3207					
C(Date) [T.2013-01-01 00:00:00]	0.1721	0.1147	1.5006	0.1335	-
0.0527 0.3968					
C(Date) [T.2013-04-01 00:00:00]	-0.0349	0.1118	-0.3119	0.7551	-
0.2539 0.1842					
C(Date) [T.2013-07-01 00:00:00]	0.1305	0.1135	1.1496	0.2503	-
0.0920 0.3530					
C(Date) [T.2013-10-01 00:00:00]	-0.0176	0.1115	-0.1577	0.8747	-
0.2360 0.2009					
C(Date) [T.2014-01-01 00:00:00]	-0.0025	0.1112	-0.0227	0.9819	-
0.2205 0.2154					
C(Date) [T.2014-04-01 00:00:00]	0.1265	0.1129	1.1211	0.2622	-
0.0947 0.3477					
C(Date) [T.2014-07-01 00:00:00]	0.1768	0.1145	1.5441	0.1226	-
0.0476 0.4012					
C(Date) [T.2014-10-01 00:00:00]	0.3338	0.1234	2.7046	0.0068	0.0919
C(Date) [T.2015-01-01 00:00:00]	0.1492	0.1140	1.3091	0.1905	-
0.0742 0.3727					

- $\hat{\text{Lag_MF_FireSale}}$ est la valeur prédite des variations de prix en utilisant Lag_MF_FireSale.
- représente les effets fixes temporels.
- est le terme d'erreur.

Résultats de la première étape – Validité de l'instrument

- Statistique F obtenue : 70.41
- Seuil critique : 10 - L'instrument est valide, car la statistique F dépasse largement le seuil recommandé de 10. Cela signifie que Lag_MF_FireSale est un bon prédicteur des variations de prix passées.

Résultats de l'estimation IV-2SLS

Statistiques générales :

- Nombre d'observations : 248 912
- R^2 ajusté : -0.0625
- Statistique F globale : 1.8099
- p-value (F-test) : 0.0000

Interprétation du R^2 : Le R^2 négatif signifie que le modèle IV-2SLS n'explique pas bien la variation de l'investissement. Cela peut être dû à une mauvaise spécification du modèle ou à une relation faible entre les variables instrumentées et l'investissement.

Interprétation des coefficients -Intercept : -0.0884 (non significatif)

- Effets fixes temporels : Certains coefficients des effets fixes temporels sont significatifs (ex. 1992-07-01 avec p-value = 0.0003), ce qui indique que certaines périodes influencent l'investissement.

Effet de Lag_MF_FireSale :

- La variable lag_log_price_change est endogène et a été instrumentée.
- L'instrument Lag_MF_FireSale a permis d'obtenir une nouvelle estimation de cet effet.
- L'interprétation économique reste cependant limitée en raison du R^2 négatif et de la faible puissance explicative du modèle.

Conclusion

Validation de l'instrument La statistique F élevée (70.41) confirme que Lag_MF_FireSale est un instrument pertinent pour expliquer les variations de prix passées.

Mauvaise explication des variations d'investissement

- Le R^2 négatif et la faible puissance explicative du modèle montrent que l'investissement des entreprises ne semble pas être influencé par les variations de prix non informées.

- Les effets fixes temporels montrent que certains chocs macroéconomiques influencent l'investissement, mais l'effet spécifique des variations de prix reste incertain.

Améliorations possibles :

- Explorer d'autres instruments pour mieux capter les variations de prix non informées.
- Inclure des effets fixes par entreprise pour mieux contrôler les facteurs spécifiques aux entreprises.
- Tester des interactions avec d'autres variables financières pour mieux comprendre la relation entre prix et investissement.

Réponses aux Questions Finales avec Analyse Approfondie

1. Pertinence de l'Approche par Variables Instrumentales (IV) avec Lag_MF_FireSale

Contexte et Justification de l'Approche IV

Dans l'analyse économétrique, une estimation par moindres carrés ordinaires (OLS) peut être biaisée lorsque la variable explicative principale est endogène, c'est-à-dire corrélée avec l'erreur résiduelle. Dans ce cas, l'inférence statistique est faussée, et les coefficients estimés ne peuvent pas être interprétés comme des effets causaux.

L'utilisation des variables instrumentales (IV) permet de pallier ce problème en remplaçant la variable explicative potentiellement endogène (`lag_log_price_change`) par une version prévisible issue d'un instrument externe (`Lag_MF_FireSale`).

Nous considérons ici `Lag_MF_FireSale` (les ventes de feu des fonds mutuels) comme un instrument potentiel des variations de prix passées pour estimer leur impact sur l'investissement.

Conditions d'un bon instrument

Un instrument doit satisfaire deux conditions clés :

1. Condition de pertinence :

L'instrument doit être corrélé avec la variable instrumentée (`lag_log_price_change`).

$$\text{Cov}(\text{Lag_MF_FireSale}, \text{lag_log_price_change}) \neq 0$$

1. Condition d'exclusion :

L'instrument ne doit pas être corrélé avec l'erreur résiduelle du modèle structurel, sauf via la variable instrumentée.

$$\text{Cov}(\text{Lag_MF_FireSale}, \text{error}) = 0$$

Autrement dit, $Lag_MF_FireSale$ ne doit pas influencer directement l'investissement des entreprises, mais uniquement par son impact sur $lag_log_price_change$.

Tests de validité de l'instrument

Pour vérifier ces conditions, nous devons effectuer plusieurs tests :

- Test de pertinence (Statistique F de la première étape) : Si la statistique F de la régression de $lag_log_price_change$ sur $Lag_MF_FireSale$ est supérieure à 10, alors l'instrument est considéré comme fort (Staiger & Stock, 1997). $F = (R^2 / (1 - R^2)) / (k - 1)$ Dans notre cas, la statistique F = 70.41, bien supérieure à 10, ce qui valide la pertinence de l'instrument.
- Test d'exogénéité (Test de Sargan-Hansen) : Ce test permet de vérifier si $Lag_MF_FireSale$ est réellement exogène.
 - 0: L'instrument est valide (corrélation avec l'erreur résiduelle nulle).
 - 1: L'instrument est corrélé avec l'erreur, et donc non valide.
 - Une p-value élevée (> 0.05) signifie que nous ne rejetons pas 0, confirmant la validité de l'instrument.

Limites de l'approche IV avec $Lag_MF_FireSale$

Malgré sa pertinence, $Lag_MF_FireSale$ comme instrument présente certaines limites :

- Possibilité d'exclusion violée :
 - Si les ventes de feu des fonds mutuels sont corrélées avec d'autres facteurs influençant directement l'investissement (comme des chocs de liquidité globaux), l'instrument pourrait être biaisé.
- Effets dynamiques ignorés :
 - L'impact des ventes de feu sur les prix peut dépendre de facteurs macroéconomiques, affectant directement l'investissement.
- Risque de faiblesse de l'instrument dans certaines périodes :
 - Même si notre test de pertinence montre une forte corrélation, la force de l'instrument peut varier selon les périodes économiques.

Conclusion :

L'instrument $Lag_MF_FireSale$ est pertinent et valide d'après nos tests. Il reste des risques de violation de la condition d'exclusion, pouvant biaiser les résultats si d'autres chocs affectent l'investissement.

2. Estimation de l'Impact des Changements de Prix Non Informés sur l'Investissement

Modèle IV-2SLS Estimé

L'estimation par deux étapes IV (2SLS) permet d'obtenir un effet causal des variations de prix non informées sur l'investissement.

Première étape (Régression des variations de prix sur l'instrument) :
$$\Delta \log P_{it} = \alpha + \beta \Delta \log P_{it}^{(1)} + \epsilon_{it}$$

Où β est le coefficient reliant $Lag_MF_FireSale$ aux variations de prix, et ϵ_{it} l'erreur résiduelle.

Deuxième étape (Estimation de l'effet des variations de prix instrumentées sur l'investissement) :
$$\hat{\Delta \text{Investissement}} = \alpha + \beta \hat{\Delta \text{Prix}} + \epsilon$$
 Où $\hat{\Delta \text{Prix}}$ est la version prédite des variations de prix utilisant l'instrument Lag_MF_FireSale.

Résultats de la Régression IV-2SLS

- Nombre d'observations : 248 912
- R^2 ajusté : -0.0625
- Statistique F : 1.8099
- p-value (F-test) : 0.0000
- Effet estimé de $\hat{\Delta \text{Prix}}$ sur l'investissement : Non significatif

Interprétation :

1. Faible pouvoir explicatif

- Le R^2 négatif indique que le modèle instrumenté explique encore moins de variance que la simple moyenne de l'échantillon.
- Cela suggère que l'instrument n'améliore pas l'explication des variations d'investissement par rapport aux modèles sans instrument.

1. Effet non significatif

- p-value élevée (> 0.05) pour le coefficient de $\hat{\Delta \text{Prix}}$, ce qui rejette l'hypothèse d'un effet significatif des variations de prix instrumentées sur l'investissement.

1. Problèmes possibles dans l'approche IV

- Lag_MF_FireSale pourrait ne pas être un instrument parfaitement valide, car il pourrait influencer directement l'investissement via d'autres canaux (ex. choc de liquidité).
- La dynamique d'ajustement des prix peut nécessiter un modèle plus complexe pour capturer les effets retardés.

Conclusion et Implications

Points positifs :

- L'instrument Lag_MF_FireSale est valide selon les tests de pertinence (Statistique F > 10).
- L'approche IV permet de contrôler l'endogénéité des variations de prix et leur impact sur l'investissement.

Limites et Problèmes :

- R^2 négatif : montre que la relation prix-investissement est mal expliquée, même après instrumentation.
- Effet non significatif : pas de preuve robuste que les variations de prix non informées influencent l'investissement.

2.5 / 4

- Problème possible de violation de la condition d'exclusion : `Lag_MF_Fire-Sale` pourrait capturer des facteurs économiques influençant directement l'investissement.

Recommandations pour Approfondir l'Analyse

1. Explorer d'autres instruments
 - Tester des instruments alternatifs liés aux flux de capitaux institutionnels ou chocs de liquidité exogènes.
1. Améliorer le modèle
 - Ajouter des effets fixes par entreprise pour capturer des spécificités sectorielles.
 - Introduire des variables interactives avec la liquidité des entreprises.
1. Tester des approches alternatives
 - Approche différence en différences (DiD) pour observer des effets causaux sur l'investissement après des chocs de marché.