

UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Scuola di Economia e Statistica

SCIENZE STATISTICHE ED ECONOMICHE

Corso di Economia e Dinamica Industriale



PERFORMANCE E INDICI DI CONCENTRAZIONE NEL  
SETTORE TRASPORTI ITALIANO

Genellini Gaia

Piazza Andrea Morena

Rausa Federico

Anno Accademico 2024/2025

## Indice

1.	INTRODUZIONE.....	1
2.	REVISIONE DELLA LETTERATURA.....	2
3.	METODOLOGIA .....	4
	3.1 <i>Indici di concentrazione e di performance</i> .....	4
	3.2 <i>Valutazione della relazione fra concentrazione e performance</i> .....	6
	3.3 <i>Test Diagnostici</i> .....	8
4.	DATI .....	9
	4.1 <i>Descrizione dati</i> .....	9
	4.2 <i>Statistiche descrittive</i> .....	11
	4.3 <i>Grafici</i> .....	12
5.	RISULTATI.....	17
	5.1 <i>Matrici di correlazione</i> .....	17
	5.2 <i>Regressione lineare</i> .....	18
	5.3 <i>Modello a effetti fissi con e senza variabili di controllo</i> .....	19
	5.4 <i>Test diagnostici</i> .....	22
6.	CONCLUSIONI.....	24
7.	BIBLIOGRAFIA E SITOGRAFIA .....	26

# 1. INTRODUZIONE

Il settore dei trasporti riveste un ruolo cruciale per l'economia italiana, fungendo da pilastro essenziale per la connessione tra regioni, imprese e mercati internazionali e contribuendo significativamente al PIL nazionale. Negli ultimi anni, il mercato ha subito grandi trasformazioni, determinate anche dalla pandemia di Covid-19, che hanno avuto un impatto non indifferente sulle performance delle imprese, sia in termini di redditività che di stabilità finanziaria, e hanno messo alla prova la capacità delle imprese di adattarsi a contesti di crisi.

L'obiettivo di questo studio è analizzare eventuali relazioni esistenti tra gli indici di concentrazione del mercato (quali  $CR_4$ , HHI e il coefficiente di Gini) e i principali indicatori di performance aziendale (ROS, ROA, ROE, EBITDA/Vendite), con un focus sul settore dei trasporti italiano. Si vuole comprendere se e come la struttura del mercato e la distribuzione delle quote influenzino le capacità delle imprese di generare profitti e mantenere la propria solidità finanziaria.

Sarà necessario definire come prima cosa la struttura del mercato, selezionare i settori di interesse ed esplorare la concentrazione delle imprese che lo compongono. Successivamente sarà possibile adottare una metodologia basata sull'analisi quantitativa dei dati, attraverso una matrice di correlazione che consenta di evidenziare le relazioni più rilevanti tra le variabili di interesse.

Successivamente si procederà con la stima di vari modelli quali i modelli ad effetti fissi con e senza variabili di controllo.

I risultati ottenuti permetteranno di delineare un quadro chiaro delle dinamiche settoriali, stabilendo il grado di concentrazione dei trasporti italiani e quanto e come esso possa influire sulle performance delle aziende.

## 2. REVISIONE DELLA LETTERATURA

L'ipotesi dell'esistenza di una relazione tra gli indici di concentrazione di un mercato e gli indici di performance di bilancio delle imprese operanti in un determinato settore è stata proposta e studiata da diversi autori. Le pubblicazioni reperite su questo tema sono datate prevalentemente nei decenni 1970-1980 e 2010-2020.

Smirlock, imitando Weiss (1974, p. 225-26), e studiando il settore bancario, utilizza i t-test sui coefficienti della regressione lineare aventi come variabile dipendente un dato indice di performance (Weiss suggerisce di adottare il ROI o il ROA) e come regressori diversi indici di bilancio, di mercato o di concentrazione (tra cui il CR, il market share MS, il loro prodotto, che chiama MSCR, e una serie di ulteriori variabili di controllo alle quali precedenti studi attribuivano un significativo impatto sulla profittabilità della banca).

Nei suoi studi Smirlock dimostra che, fornito il market share, la concentrazione di mercato non contribuisce a migliorare la significatività del modello, almeno nel settore bancario. Pochi anni dopo, nel 1981, Phillips dimostrò la stessa tesi, e anche Kwoka nel 1971.

I modelli di regressione adottati spesso vengono applicati su dati panel. Nasce quindi l'esigenza di adottare modelli più robusti dei classici OLS, e in particolare molti autori, fra cui Hedges et al., prediligono i modelli a effetti fissi e a effetti random.

Strickland e Weiss studiano la relazione tra concentrazione e pubblicità nel 1976 e propongono un modello più consistente dei minimi quadrati, detto a equazioni simultanee, adatto a modellare fenomeni che si influenzano a vicenda (come appunto il potere di mercato e gli investimenti in pubblicità) partendo dalla condizione di Dorfman Steiner.

Lee utilizza invece, per la prima volta contemporaneamente, modelli a effetti fissi, statici e dinamici, per lo studio della relazione tra la crescita della dimensione aziendale e i profitti delle imprese della Corea del Sud (senza porre filtri alle tipologie di settore).

Il suo studio era stato precedentemente affrontato da Cowling, Coad et al., Goddard et al., Jang e Park tra il 2004 e il 2011: tutti loro avevano utilizzato un modello OLS, 2SLS, GMM o LAD per stimare come variabile dipendente indici di profittabilità come VA, ROS, ROE, profitti, GOS, o OS, utilizzando come unici regressori il fatturato, il numero di dipendenti o l'attivo.

Di contro, Salinger, molti anni prima, nello studio del settore manifatturiero, considerava il tobin-q come l'indice di performance più legato agli indici di concentrazione di mercato  $CR_4$ ,  $CR_8$  e HHI. In genere si vede che nella misura della concentrazione di mercato vengono adottati prevalentemente gli indici CR e HHI e quasi mai l'indice di Gini. Tuttavia, Reynolds-Feighan, nello studio del settore dei trasporti aerei US, utilizza esclusivamente Gini; perciò, potrebbe essere ragionevole includere questa misura nelle analisi del settore dei trasporti. Curry e George (1983) non utilizzano Gini, ma l'indice di Entropia, che può considerarsi una misura analoga, e includono congiuntamente CR e HHI.

Molti ricercatori, oltre ad adottare metodi di regressione o di regressione robusta, si avvalgono anche dei basilari indici di correlazione bivariata, come Pearson, Spearman e Kendall. Pearson è il più comune, perché misura una correlazione di tipo lineare e può essere affiancato a modelli di regressione. Specularmente, Spearman coglie anche relazioni non lineari fra le variabili, e per questo non viene adottato frequentemente nelle analisi economiche, nelle quali occupano il primo posto quasi sempre modelli lineari per la regressione. Kendall è il meno usato, in quanto comporta un costo computazionale spesso insostenibile.

Ornstein (1972) utilizza OLS e coefficiente di Pearson, nonché dei test sulla correlazione, per misurare la relazione tra i rapporti di concentrazione e degli indici di profittabilità.

Černohorský e Prokop (2016), studiando il settore bancario di Austria, Belgio e Repubblica Ceca, utilizzano gli indici di correlazione di Pearson e di Spearman per identificare la presenza di relazioni non lineari tra gli indici di profittabilità e di concentrazione.

Uscendo dal contesto economico, la misura e il test della correlazione fra due variabili è stata ampiamente discussa in letteratura.

Chok mette a confronto gli indici di correlazione di Pearson e quelli di Spearman, per diverse tipologie di variabili continue. Counsell e Cribbie evidenziano come un'analisi della relazione tra due variabili basata sui coefficienti di correlazione e una basata sui coefficienti di regressione possano portare a risultati differenti.

Asuero et al. consigliano un utilizzo congiunto dei due metodi, ma raccomandano la trasformazione di Fisher sul coefficiente di correlazione di Pearson per garantirne la normalità. Diversi degli autori fin qui citati fanno ricorso a test-t o chi-quadro sul coefficiente di Pearson per ricavare un p-value sull'ipotesi nulla di indipendenza fra due variabili.

Altri autori, come DiCiccio e Romano., utilizzano test non parametrici sui coefficienti di correlazione, come il bootstrap (per simularne la distribuzione, estraendo solo alcune coppie), o i test di permutazione (cambiando l'ordine delle osservazioni della seconda variabile, e verificando che la correlazione associata sia significativamente inferiore rispetto a quella effettiva). Essi propongono l'adozione dei test di permutazione anche sui coefficienti di regressione. Krzanowski adotta i permutation test su matrici di correlazione. Chatzipantsiu et al. riescono a ridurre notevolmente il costo computazionale di questi metodi nell'utilizzo di R.

Diversi autori prediligono il test di Hausman per decidere se adottare un modello a effetti fissi o un modello a effetti random. Alcuni mettono in discussione questo metodo, come Sheytanova, il quale, utilizzando il bootstrap su tale test nelle sue analisi, produce un sovrarigetto dell'ipotesi nulla.

### 3. METODOLOGIA

L'analisi è stata svolta utilizzando il software RStudio sul quale è stato caricato il dataset contenente le variabili di interesse (colonne) e tutte le imprese del settore dei trasporti (righe).

Per procedere è stato necessario creare una lista di 15 dataset, segmentazione che è stata generata selezionando le prime tre cifre del codice NACE Rev. 2 e che ha portato alla determinazione di 15 sottosettori. Successivamente ognuno di essi è stato trasformato nel formato *Data Panel*, combinando le informazioni relative alle caratteristiche di N individui nello stesso istante temporale con quelle rilevate per gli stessi individui in T diversi periodi di tempo. Nei modelli di tipo panel i dati disponibili hanno entrambe le caratteristiche di:

- Dati *Cross Section*, che per un dato istante osservano le caratteristiche di più individui;
- Dati *Time Series*, che rilevano le diverse caratteristiche in diversi istanti temporali per un dato collettivo di individui.

Nel caso in esame,  $T = 2015 \rightarrow 2023$  e  $N = 1 \rightarrow 94554$ .

#### 3.1 Indici di concentrazione e di performance

Dopo aver preparato i dataset nel formato opportuno, sono stati calcolati gli indici di concentrazione di mercato e quelli di performance per ognuno dei 15 sottosettori e per ogni anno a disposizione all'interno di ogni sottosettore.

Gli indici di concentrazione presi in esame sono:

- Four-firm concentration ratio ( $CR_4$ ):

è pari alla somma delle quote di mercato dei quattro operatori con la quota di mercato più elevata. L'indice  $CR_4$  è compreso tra 0 (caso di concorrenza perfetta) e 100 (monopolio).

$$CR_4 = 100 * \frac{(Ricavi_A + Ricavi_B + Ricavi_C + Ricavi_D)}{Ricavi_S}$$

con A, B, C e D che indicano le quattro società più grandi del settore S;

- Indice di Herfindahl-Hirschman (HHI):

è pari alla somma del quadrato delle quote di mercato di tutte le imprese all'interno di un mercato, il tutto moltiplicato per 10000. Valori bassi dell'indice (inferiori ai 1000 punti) indicano un settore molto competitivo. Se l'indice è compreso tra i 1000 e i 1800

punti il comparto viene definito moderatamente competitivo. Sopra i 1800 punti nel settore c'è poca competizione. Un indice pari a 10000 punti indica un monopolio.

$$HHI = 10000 * \sum_{i=1}^N s_i^2$$

con  $s_i = \frac{Ricavi_i}{Ricavi_S}$ , che indica la quota dell'i-esima impresa nel settore S.

- Indice di Gini (GINI):

è una misura della disuguaglianza di una distribuzione. È spesso usato come indice di concentrazione per misurare la disuguaglianza nella distribuzione del reddito o anche della ricchezza. È un numero compreso tra 0 ed 1. Valori bassi del coefficiente indicano una distribuzione abbastanza omogenea, con il valore 0 che corrisponde alla pura equidistribuzione (Concorrenza Perfetta); valori alti del coefficiente indicano una distribuzione più diseguale, con il valore 1 che corrisponde alla massima concentrazione (monopolio).

Gli indicatori di performance analizzati sono:

- Redditività del capitale proprio (ROE – Return on Equity):

è l'indicatore che viene utilizzato per verificare il tasso di remunerazione del capitale di rischio, ovvero quanto rende il capitale conferito all'azienda dai soci. Può essere considerato la valutazione di come il management sia riuscito a gestire i mezzi propri per aumentare gli utili aziendali.

Si calcola come:

$$ROE = Utile Netto / Capitale Proprio * 100$$

- Redditività delle vendite (ROS-Return on Sale):

è impiegato per valutare l'efficienza operativa di un'azienda, ovvero per capire quanto sia in grado di trasformare le vendite in profitti. In particolare, fornisce informazioni su quanto profitto viene prodotto per ogni euro incassato dalle vendite.

La sua formula è data da:

$$ROS = Utile operativo (EBIT) / Vendite nette * 100$$

- EBITDA/Vendite:

è l'indicatore che esprime quanto reddito operativo è in grado di generare l'impresa per unità di fatturato. Un buon valore di questo indice si ha quando è maggiore al 10%;

- Redditività del totale attivo (ROA-Return on Assets):

è un indice di bilancio che misura la redditività di un'impresa in relazione alle risorse utilizzate per svolgere la propria attività economica.

Viene calcolato come segue:

$$ROA = Utile\ Netto / Totale\ Attivo * 100$$

In aggiunta a questi indicatori sono state utilizzate ulteriori variabili che forniscono informazioni circa le performance aziendali, quali:

- Debt Equity Ratio (rapporto di indebitamento):

evidenzia il peso del debito finanziario rispetto alle risorse durevoli dell'impresa: ossia il capitale, le riserve e gli utili non distribuiti. Maggiore è il peso dei debiti finanziari rispetto al patrimonio netto dell'azienda, maggiore è il rischio della struttura finanziaria;

- Totale Attività:

è la somma degli investimenti, rimanenze, crediti e liquidità;

- Totale Debiti:

è il totale di tutti i debiti accolti nel passivo dello stato patrimoniale, siano essi a breve o a medio-lungo termine;

- Capitale sociale:

rappresenta il valore delle somme e dei beni conferiti dai soci, a titolo di capitale di rischio, all'atto della costituzione dell'impresa.

### *3.2 Valutazione della relazione fra concentrazione e performance*

Per analizzare la possibile relazione esistente fra la concentrazione di mercato e la performance sono stati presi in considerazione tre approcci differenti:

1. Correlazione di Pearson e di Spearman:

- la correlazione di Pearson calcola l'effetto della variazione di una variabile al variare dell'altra; esprime un'eventuale relazione di linearità tra esse. Assume un valore compreso tra +1 e -1, dove +1 corrisponde alla perfetta correlazione lineare positiva, 0 corrisponde a un'assenza di correlazione lineare e -1 corrisponde alla perfetta correlazione lineare negativa. Mediante il comando



cor() in RStudio viene generata una matrice di correlazione fra gli indici che sono stati descritti precedentemente;

- la correlazione di Spearman riesce a cogliere la correlazione tra due variabili che presentano una relazione funzionale quadratica o di polinomio n o comunque non lineare.

Se il coefficiente di Pearson è alto, e il coefficiente di Spearman è alto, allora le due variabili sono linearmente correlate (vince Pearson).

Se sono entrambe basse, non c'è alcuna correlazione, né lineare né non lineare.

Se Pearson è basso, e Spearman è alto, la correlazione è presente ed è non lineare.

Il caso Pearson alto e Spearman basso è semplicemente impossibile che si verifichi;

## 2. Regressione lineare:

tramite il comando lm() in RStudio si stima la retta di regressioni lineare ed essa viene confrontata con la distribuzione dei dati per evidenziare un possibile andamento lineare tra concentrazione e performance;

## 3. Modelli per dati panel a effetti fissi (dummy variable model):

fa parte della famiglia dei modelli lineari di regressione applicati in un contesto di dati longitudinali. In questa analisi vengono utilizzati i modelli a effetti fissi, ovvero si considerano le eterogeneità individuali costanti nel tempo e non casuali: equivale ad ipotizzare che le differenze tra individui sono catturate da differenze nella costante della regressione lineare.

### ➤ Modelli senza variabili di controllo:

vengono utilizzate come variabili indipendenti gli indici di concentrazione del mercato ( $CR_4$ , HHI, e Gini) e come variabili dipendenti gli indicatori di performance (Redditività delle vendite – ROS, Redditività del capitale proprio – ROE, Redditività del totale attivo – ROA ed EBITDA/Vendite);

### ➤ Modelli con variabili di controllo:

vengono implementate i medesimi modelli del caso precedente e vi si aggiungono anche delle variabili di controllo finanziarie (Debt Equity ratio, il logaritmo del Totale Attività, il logaritmo del Totale Debiti e il logaritmo del Capitale Sociale).

### 3.3 Test Diagnostici

Per completare l'analisi vengono effettuati dei test per valutare la validità dei modelli.

I test utilizzati sono:

- Test di Hausman:

permette di capire se è più appropriato l'utilizzo di un modello a effetti fissi o di uno a effetti random;

$H_0$ : assenza di correlazione tra effetti individuali e regressori

$H_1$ : presenza di correlazione tra effetti individuali e regressori

Sotto l'ipotesi nulla gli stimatori Random Effect e Fixed Effect sono consistenti e le stime Random Effect sono più efficienti, sotto l'ipotesi alternativa Fixed Effect è consistente mentre Random Effect è inconsistente.

Ne segue che, se vi è evidenza empirica a favore dell'ipotesi nulla, si ha assenza di correlazione ed è quindi preferibile utilizzare il modello a effetti random poiché più efficiente (anche se entrambi i modelli hanno stime consistenti). Se, al contrario, l'evidenza empirica suggerisce la presenza di correlazione, è preferibile l'utilizzo di un modello a effetti fissi;

- Test di Breusch-Pagan per effetti individuali random:

permette di valutare se è più appropriato l'utilizzo di un modello pooling semplice (senza effetti individuali) rispetto ad un modello con effetti individuali casuali.

Se i dati suggeriscono il rifiuto dell'ipotesi nulla significa che gli effetti casuali individuali non sono statisticamente significativi quindi è preferibile utilizzare il modello pooling semplice;

- Test di Breusch-Pagan per l'eteroschedasticità:

è il test utilizzato per verificare la presenza di eteroschedasticità all'interno del modello lineare di regressione. L'ipotesi nulla del test è l'assenza di eteroschedasticità;

- Test di Wooldridge:

permette di verificare se vi è presenza di autocorrelazione nel modello. Se l'evidenza empirica suggerisce il rifiuto dell'ipotesi nulla, significa che vi è autocorrelazione.

## 4. DATI

### 4.1 Descrizione dati

Per la realizzazione di quest'analisi sono stati utilizzati i dati dei bilanci delle società operanti nel settore dei trasporti forniti dalla banca dati AIDA (Analisi Informatizzata Delle Aziende), realizzata e distribuita da Bureau van Dijk S.p.A., che fornisce informazioni economico-finanziarie sulle aziende di capitale italiane attive e fallite.

Per identificare le imprese da includere nel settore di interesse, si fa riferimento alla classificazione standard NACE Rev. 2, adottata dall'Unione Europea per la nomenclatura statistica delle attività economiche, così da rendere possibili eventuali comparazioni con altri Paesi europei. La sezione scelta è la H – Trasporto e Magazzinaggio e i dati relativi sono analizzati per gruppo (3 cifre), ovvero a livello della sottocategoria della divisione (e.g. 511 Trasporto aereo di passeggeri).

Selezionata la granularità del dato ritenuta adeguata, è necessario identificare le variabili rilevanti ai fini dell'analisi, come indicatori di performance aziendali, e variabili di controllo che consentano di comprendere cosa influenzi le prestazioni dell'impresa. Le variabili selezionate, comprese quelle riguardanti le informazioni anagrafiche delle aziende, sono le seguenti:

- Ragione sociale
- Provincia
- Codice fiscale
- Numero CCIAA
- Totale Attività (in migliaia di euro)
- Ricavi delle vendite (in migliaia di euro)
- Utile netto (in migliaia di euro)
- EBITDA (in migliaia di euro)
- EBITDA/Vendite
- Redditività delle vendite (ROS)
- Redditività del totale attivo (ROA)
- Redditività del capitale proprio (ROE)
- Debt/Equity ratio
- Dipendenti
- Capitale sociale (in migliaia di euro)
- Totale debiti (in migliaia di euro)
- Codice NACE Rev. 2
- Descrizione del codice
- Data di chiusura dell'ultimo bilancio
- Indicatore di indipendenza (BvD)

- Number of companies in corporate group
- Numero azionisti registrati
- Number of participate registrate

I dati così selezionati vengono estratti da AIDA per un intervallo di tempo dal 2015 al 2023 e scaricati in formato Excel, ottenendo un data frame di 94556 aziende e 120 variabili, che verrà poi elaborato sul software RStudio.

Per facilitare l'analisi, si procede con il processing dei dati.

La prima operazione consiste nella rimozione dai nomi delle colonne di eventuali caratteri speciali. Successivamente, viene aggiunta una nuova variabile contenente le prime tre cifre del codice NACE, variabile che viene utilizzata per ordinare il dataset. A questo punto, si sostituisce nei dati la dicitura "n.d." con "NA" e si rimuovono le ultime due osservazioni che risultano essere prive di informazioni.

Il dataset comprende ora 94554 aziende e viene suddiviso in una lista di 15 data frame sulla base del codice NACE a 3 cifre. Questo passaggio consente di trasformare ogni lista in formata panel data e facilitare le analisi temporali e longitudinali sui dati.

A questo punto, si procede con la trasformazione delle variabili nei formati adeguati (e.g. la variabile ROA viene convertita da *character* a *numeric*) per l'analisi.

Una volta processato il dataset, è possibile procedere con il calcolo degli indici di concentrazione ( $CR_4$ , HHI e Gini) basati sulla variabile Ricavi delle vendite per tutti i 15 dataset in formato panel, ottenendo una lista di indici per ciascun gruppo settoriale.

Calcolati gli indici di concentrazione, si prepara un unico data frame che combina gli indici di concentrazione di ciascun settore e contiene sia le informazioni settoriali che gli anni, questi ultimi in formato *data*. Dal data frame così strutturato sarà possibile ottenere i *time series plot* degli indici.

Per quanto riguarda le variabili di performance aziendale, si procede selezionando le variabili riguardanti il ROS, il ROE, il ROA e il rapporto tra EBITDA e le vendite, suddividendo i dati per settore, e anche da essi si otterranno i grafici delle serie temporali.

Per valutare l'impatto degli indici di concentrazione sulle performance delle imprese, vengono combinate le medie annuali delle variabili di performance con gli indici di concentrazione in un unico data frame, pronto per l'analisi statistica.

Successivamente, alcune variabili monetarie vengono trasformate in logaritmo naturale per normalizzarne la distribuzione ed evidenziare relazioni proporzionali nelle analisi successive.

## 4.2 Statistiche descrittive

Vengono selezionate alcune variabili di interesse economico-finanziario per il calcolo delle statistiche descrittive, i cui risultati sono riportati nella Tabella 1. Gli indicatori selezionati offrono una panoramica delle dinamiche sia strutturali che di performance delle imprese operanti nel settore e suggeriscono un mercato caratterizzato da alta variabilità e differenze strutturali non irrilevanti.

Tabella 1: Statistiche descrittive

	Media	Deviazione_Standard	Minimo	Massimo	Mediana	Q1	Q3
CR4	62.237	31.255	7.434	99.993	74.468	30.830	88.587
HHI	2689.086	2674.730	27.891	9995.524	1650.082	352.351	4229.205
Gini	0.855	0.093	0.579	0.983	0.859	0.824	0.921
Redditività_delle_vendite_ROS	3.403	2.670	-9.778	10.603	3.455	2.380	4.853
Redditività_del_totale_attivo_ROA	-2.057	7.088	-27.184	14.042	-1.089	-6.791	2.111
Redditività_del_capitale_proprio_ROE	8.247	7.950	-25.318	31.296	9.237	3.772	14.108
EBITDA_Vendite	-1.722	17.541	-60.712	37.337	1.449	-10.434	7.001
Debt_Equity_ratio	1.100	2.894	-21.791	8.925	1.103	0.530	1.678
log_Totale_Attività	10.055	2.468	6.222	16.438	9.390	8.571	10.704
log_TOTALE_DEBITI	9.544	2.563	5.582	16.357	8.847	7.868	10.305
log_Capitale_sociale	7.485	2.382	3.007	12.003	7.124	6.338	8.867

Nel complesso, le distribuzioni delle variabili sono asimmetriche, con media e mediana distinte. Le variabili con maggiore variabilità sono CR<sub>4</sub>, con deviazione standard pari a 31.255, EBITDA/Vendite (17.541) e HHI, che è quella che presenta la maggiore variabilità tra gli indicatori selezionati, con deviazione standard di 2674.730.

Nello specifico, l'indice CR<sub>4</sub> presenta una media di 62.237 ovvero le prime 4 imprese dell'industria considerata detengono in media il 62% del mercato, suggerendo una moderata concentrazione. Il massimo di questo indice è di 99.993, indicando la presenza di sottosettori con elevata concentrazione.

L'indice HHI è in media pari a 2689.086, una concentrazione moderata-alta. L'alta variabilità è confermata da un minimo pari a 27.891 e un massimo di 9995.524, che tende alla concentrazione di monopolio.

Con una media di 0.855, l'indice di Gini mostra ancora una volta un'elevata disuguaglianza nella distribuzione.

Il ROS, con distribuzione centrata verso il lato positivo come mostrato dai quartili, varia da un minimo di -9.778, con imprese in perdita nel settore, a un massimo di 10.603, con redditività piuttosto elevata. La media positiva e pari a 3.403 mostra una redditività perlopiù positiva nel settore dei trasporti.

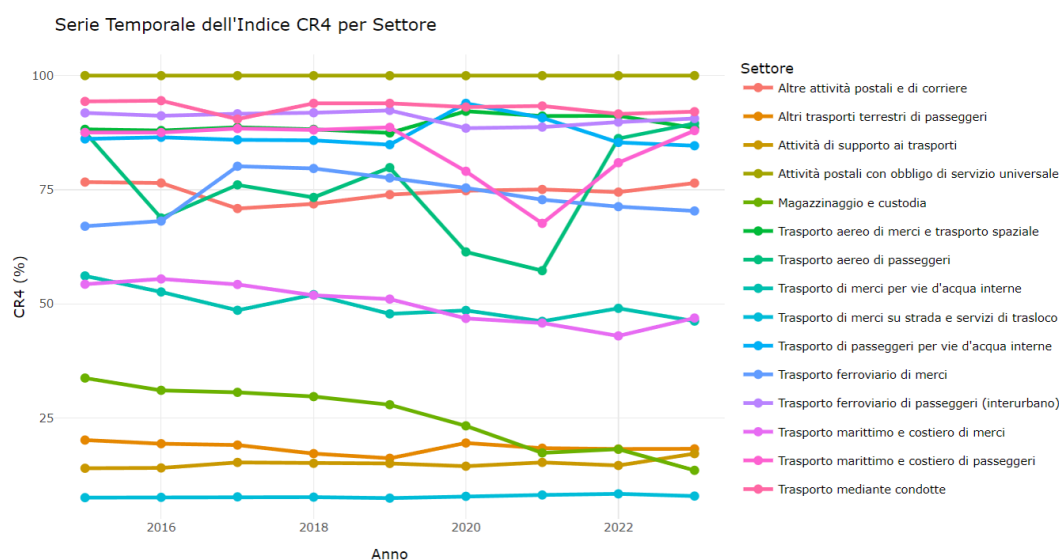
Il ROA è in media negativo, denota una perdita relativa agli attivi totali, ma la deviazione standard riflette un'alta variabilità, confermata da minimo pari a -27.184 e un massimo di 14.042, casi estremi sia negativi che positivi.

Il ROE indica una buona redditività media del capitale proprio ma con significativa variabilità, confermata da valori estremi in entrambe le direzioni.

### 4.3 Grafici

Oltre alle statistiche descrittive, può essere utile visualizzare l'andamento dei principali indicatori relativi al settore dei trasporti nel corso degli anni dal 2015 al 2023. Un'analisi congiunta dei grafici consentirà di comprendere le caratteristiche e tendenze del mercato.

Figura 1



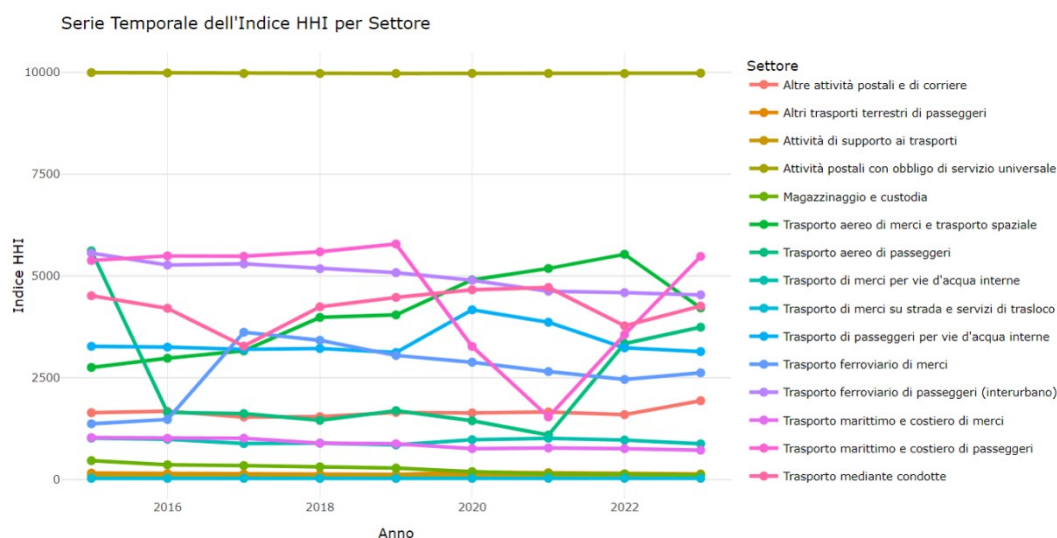
Il grafico riportato nella Figura 1 mostra l'andamento temporale dell'indice  $CR_4$ . È possibile notare che i settori “Attività postali con obbligo di servizio universale” e “Trasporto mediante condotte” mostrano valori costantemente vicini al 100%, suggerendo un mercato quasi monopolistico o fortemente oligopolistico, una stabilità che denota una struttura di mercato consolidata e dominata da poche imprese.

Tra i settori a concentrazione medio-alta, tra il 60 e 80%, rientrano “Trasporto ferroviario di passeggeri (interurbano)” e “Trasporto marittimo e costiero di passeggeri” con una concentrazione significativa ma meno estrema e variazioni limitate nel corso degli anni. I settori di questa fascia sembrano essere meno competitivi rispetto ad altri, ma non totalmente dominati da poche imprese.

Con un valore di  $CR_4$  sotto il 60% rientrano settori tra cui “Magazzinaggio e custodia”, “Trasporto di merci su strada e servizi di trasloco” e “Trasporto aereo di merci e trasporto spaziale”. La loro maggiore frammentazione indica una tendenza alla competizione.

Il settore “Trasporto aereo di passeggeri” mostra una forte variazione tra il 2020 e il 2022, con un calo seguito da una ripresa della concentrazione, andamento che potrebbe essere spiegato dalla pandemia di Covid-19, che ha avuto ripercussioni sulle dinamiche economiche.

Figura 2



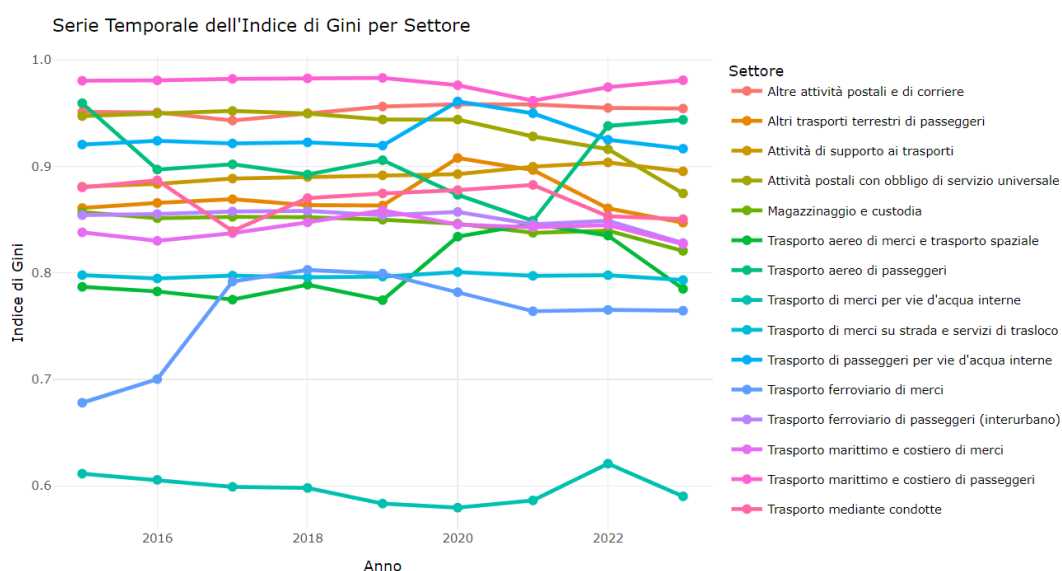
Il grafico dell'indice HHI per settore (Figura 2) sembra suggerire lo stesso andamento della concentrazione di mercato. In particolare, per quanto riguarda il “Trasporto aereo di passeggeri”, si nota una tendenza verso una concentrazione più elevata negli anni recenti, probabilmente dovuta a fusioni o riorganizzazioni del mercato.

Settori come “Altri trasporti terrestri di passeggeri” e “Attività di supporto ai trasporti” mostrano invece una relativa stabilità dell'HHI negli anni in esame, indicando un mercato abbastanza consolidato e senza grandi cambiamenti nella distribuzione delle quote.

Variazioni significative si riscontrano negli andamenti di “Trasporto marittimo e costiero di passeggeri” e “Trasporto aereo di passeggeri”, che mostrano un incremento dell'HHI negli ultimi anni, segno di una possibile maggiore concentrazione del mercato.

L'andamento dell'indice di Gini nel settore trasporti è evidenziato nella Figura 3.

Figura 3

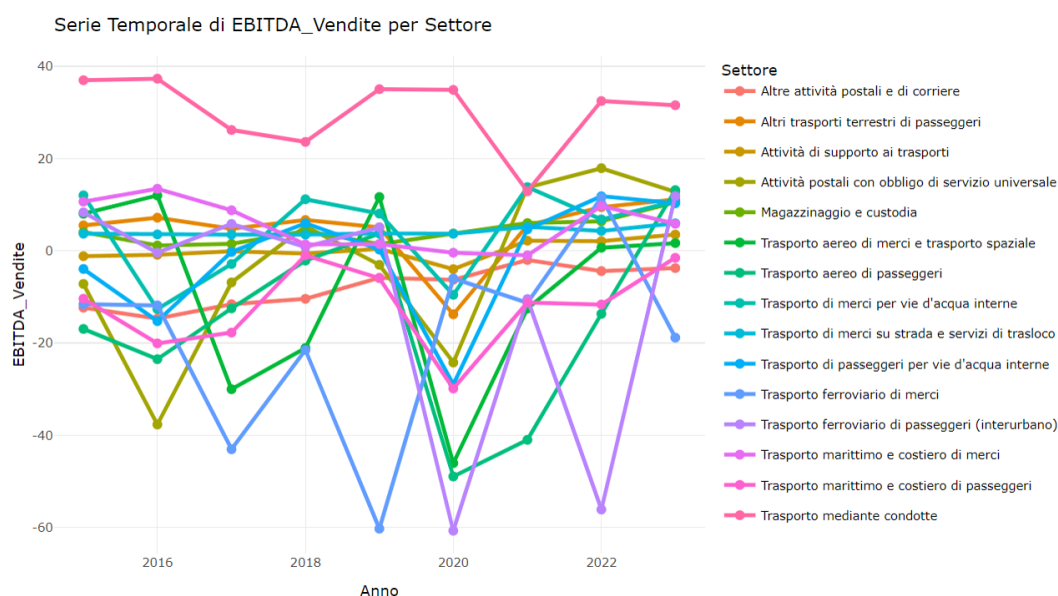


È possibile confermare l'eterogeneità della concentrazione di mercato nel settore dei trasporti. Nel contesto marittimo e postale si conferma un'alta concentrazione nel corso degli anni, mentre in altri, come il trasporto ferroviario di merci, è presente una maggiore distribuzione delle quote di mercato e un miglioramento della competitività.

Alcuni mercati stanno dunque affrontando una diversificazione, mentre altri si stabiliscono verso una concentrazione più elevata.

La Figura 4 mostra l'andamento dell'EBITDA/Vendite dei vari settori.

Figura 4



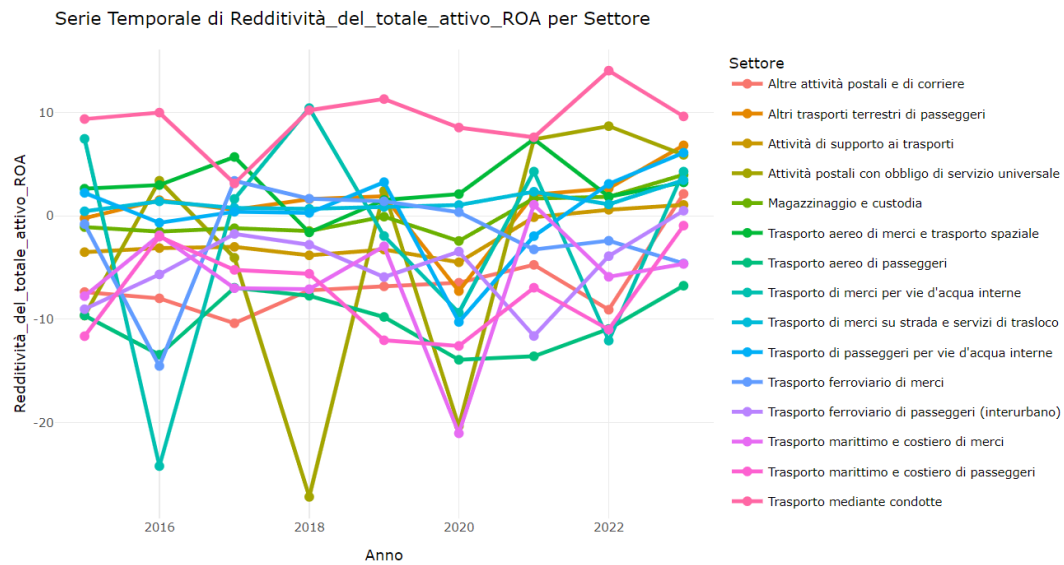
Il grafico mette in evidenza una forte variabilità nei trasporti, con alcune aree che godono di una redditività stabile (tra cui le attività postali) e altre invece in difficoltà, come i trasporti interni via acqua. In corrispondenza del 2020 sono presenti forti fluttuazioni, sintomo della sensibilità del settore alle dinamiche economiche globali. Nel periodo post-pandemia si assiste in genere a un trend positivo, in corrispondenza di una ripresa.

Nella Figura 5 è possibile analizzare l'andamento del ROA, che mette in evidenza ancora una volta le disparità nel settore.

Le attività postali e il trasporto marittimo si contraddistinguono per la loro stabilità nelle performance, mentre altre mostrano una forte vulnerabilità a shock economici e dinamiche di mercato. Le conseguenze della pandemia sono ancora chiaramente visibili, ma le tendenze di ripresa in alcuni settori suggeriscono opportunità di crescita nel periodo post-pandemia.

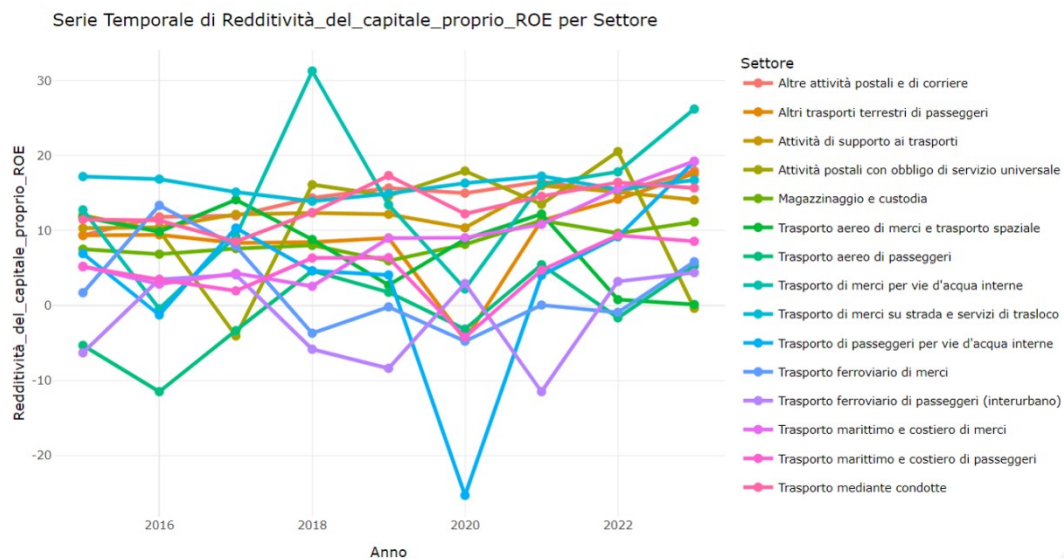


Figura 5



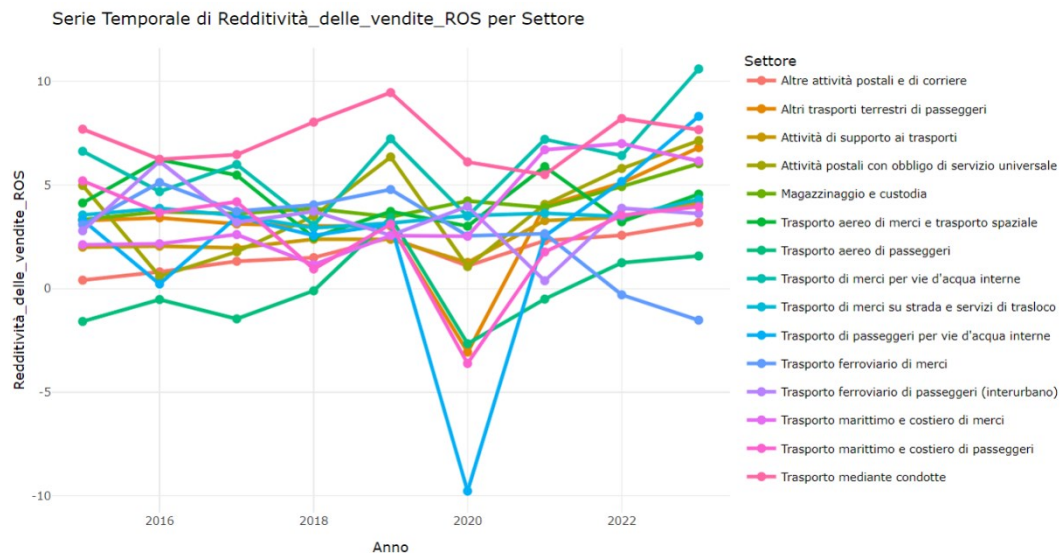
Il grafico in Figura 6, riguardante l'andamento del ROE, sottolinea le differenze strutturali tra i vari settori: se alcuni, come le attività postali e il trasporto aereo di merci, mostrano stabilità e capacità di generare valore, altri risultano destabilizzati dagli shock economici.

Figura 6



Rispetto a ROA e ROE, il ROS (in Figura 7) si concentra esclusivamente sull'efficienza operativa. Il grafico evidenzia una ripresa generale post-pandemia, con alcuni settori che hanno dimostrato maggiore capacità di adattamento rispetto ad altri.

Figura 7



## 5. RISULTATI

### 5.1 Matrici di correlazione

È possibile valutare la relazione tra gli indicatori di performance aziendali e gli indici di concentrazione tramite le matrici di correlazione riportate in Figura 8.

Figura 8.a Correlazione di Pearson

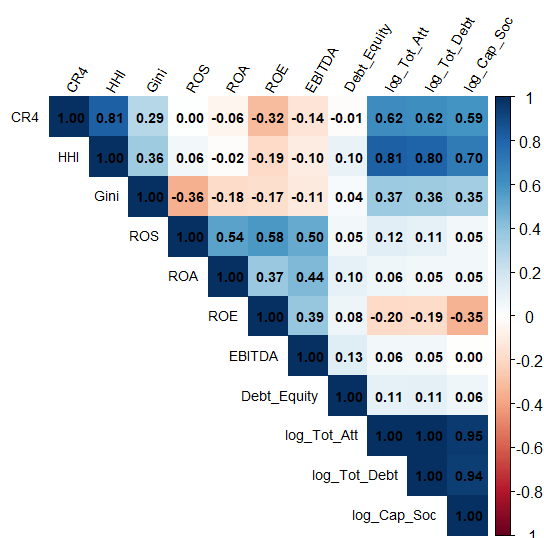
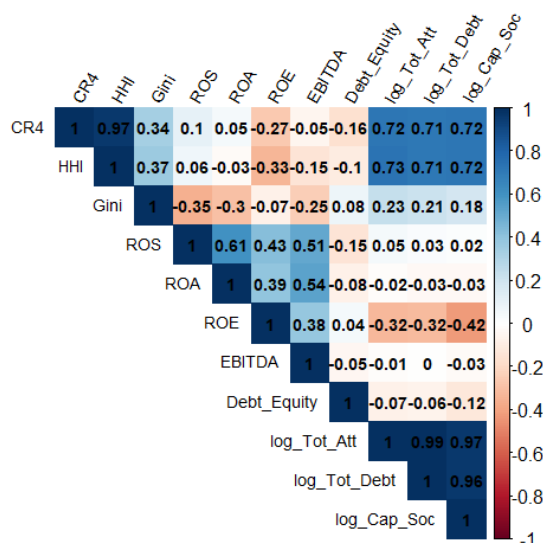


Figura 8.b Correlazione di Spearman



Risulta evidente dalla correlazione lineare e non tra HHI e CR<sub>4</sub> piuttosto elevata (0.81 e 0.97) che i due indici rappresentano coerentemente la concentrazione presente nel mercato dei trasporti. L'indice di Gini presenta correlazione con gli altri due decisamente più ridotta, cogliendo aspetti diversi della distribuzione dell'industria.

Tra gli indici di performance aziendale, il ROS sembra essere quello maggiormente correlato con gli altri, mentre l'EBITDA presenta una correlazione quasi nulla con gli altri indici di performance.

Le variabili riguardanti il logaritmo del totale dell'attività e del totale del debito sono altamente correlate, così come il logaritmo del capitale sociale è fortemente in relazione positiva con entrambe le variabili, mettendo in evidenza una coerenza tra il capitale sociale e le dimensioni aziendali.

Riguardo alla diretta correlazione tra concentrazione e performance, CR<sub>4</sub> e HHI non presentano una forte relazione diretta con ROS, ROA e ROE; piuttosto, sembrano esserci correlazioni positive tra CR<sub>4</sub> e il totale dell'attività e il capitale sociale, suggerendo che la

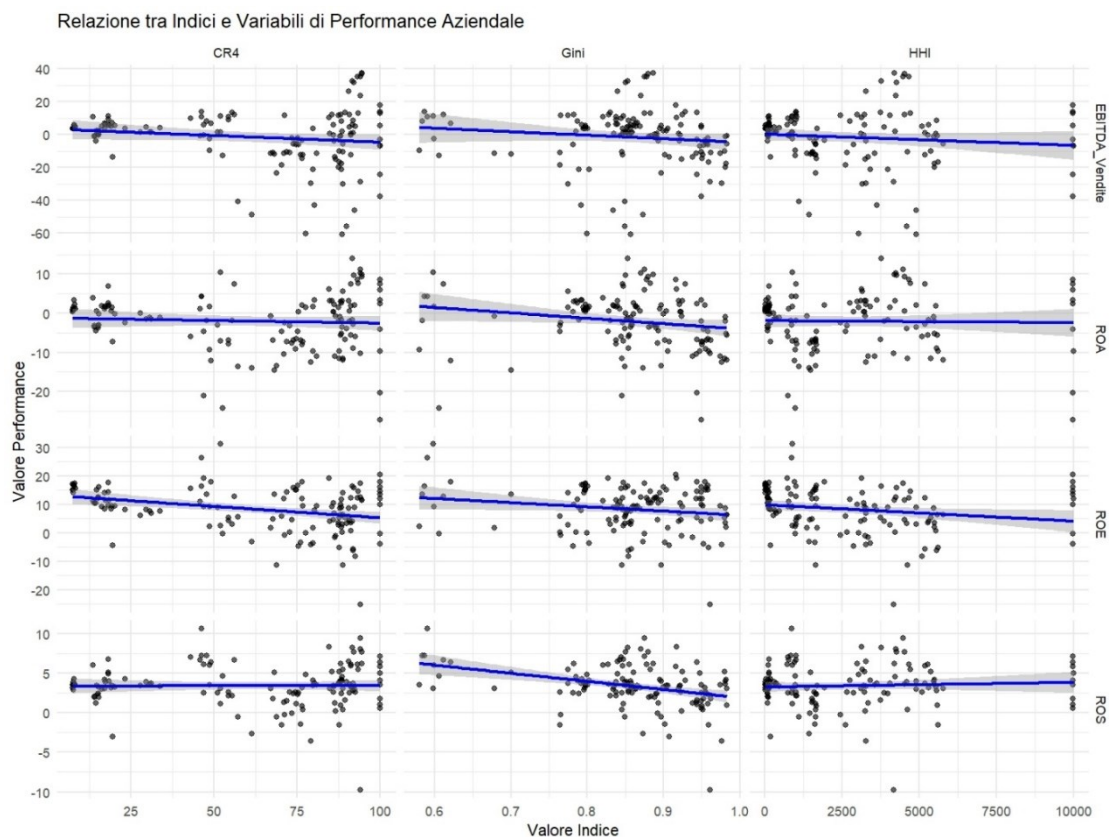
concentrazione potrebbe influire sugli aspetti patrimoniali e viceversa e che le aziende in mercati più concentrati tendano ad essere più grandi, come ci si aspetta.

Nel periodo 2015-2023, dunque, la concentrazione di mercato non ha un impatto diretto molto significativo sulla prestazione delle aziende nel settore dei trasporti. I risultati, tuttavia, potrebbero essere influenzati dalla presenza nel dataset degli anni della pandemia.

## 5.2 Regressione lineare

Un'alternativa alla matrice di correlazione è l'utilizzo di una semplice regressione lineare tra i diversi indici di concentrazione e le misure di performance (Figura 9).

Figura 9



Per quanto riguarda il  $CR_4$ , esso sembra seguire un andamento abbastanza lineare con gli indicatori EBITDA/Vendite, ROA, ROE e ROS, mentre l'indice di Gini e l'HHI si allontanano maggiormente dalla retta di regressione. In generale, si può osservare una tendenza decrescente, suggerendo che all'aumentare della concentrazione nel mercato vi sia un ridursi delle performance: questo potrebbe indicare il formarsi di un mercato meno

dinamico nel caso di settori tendenti al monopolio, nei quali si tende a trascurare l'efficienza logistica per la mancanza di competizione.

I punti sono molto dispersi nei grafici, dunque potrebbero esserci altre variabili da considerare per spiegare la relazione. È evidente l'importanza di valutare anche modelli più elaborati e adeguati al contesto panel.

### 5.3 Modello a effetti fissi con e senza variabili di controllo

Si procede inizialmente con la stima di modelli a effetti fissi senza variabili di controllo. I risultati vengono illustrati nella Figura 10 e nella Figura 11.

Figura 10

Risultati dei Modelli senza Variabili di Controllo						
Dependent variable:						
	Redditività_delle_vendite_ROS			Redditività_del_capitale_proprio_ROE		
	(1)	(2)	(3)	(4)	(5)	(6)
CR4	-0.027 (0.044)			-0.239* (0.122)		
HHI		-0.00003 (0.0003)			-0.001 (0.001)	
Gini			-22.420** (9.692)			-56.329** (27.422)
Observations	135	135	135	135	135	135
R2	0.003	0.0001	0.043	0.031	0.018	0.034
Adjusted R2	-0.123	-0.126	-0.078	-0.091	-0.105	-0.087
F Statistic (df = 1; 119)	0.373	0.008	5.351**	3.849*	2.214	4.219**
Note: *p<0.1; **p<0.05; ***p<0.01						

Figura 11

	Redditività_del_totale_attivo_ROA			EBITDA_Vendite		
	(7)	(8)	(9)	(10)	(11)	(12)
CR4	0.098 (0.112)			0.358 (0.278)		
HHI		0.001 (0.001)			0.0004 (0.002)	
Gini			-16.105 (25.362)			-116.856* (62.376)
-----						
Observations	135	135	135	135	135	135
R2	0.006	0.008	0.003	0.014	0.0003	0.029
Adjusted R2	-0.119	-0.117	-0.122	-0.111	-0.126	-0.094
F Statistic (df = 1; 119)	0.769	0.962	0.403	1.655	0.039	3.510*
=====						
Note:						*p<0.1; **p<0.05; ***p<0.01

Dall'analisi dei p-value emerge che gli unici modelli con esplicativa significativa al 5% sono quelli in cui l'indice di Gini spiega in un caso il ROS e nell'altro il ROE, in entrambi i casi con coefficiente negativo, confermando la relazione intuibile dai grafici della regressione lineare.

L'indice di Gini risulta significativo al 10% come esplicativa nel modello di EBITDA/Vendite, così come l'indice CR<sub>4</sub> nel modello che spiega il ROE. Anche in questo caso i coefficienti delle esplicative sono negativi.

Il coefficiente di determinazione R<sup>2</sup> risulta ridotto per tutti i modelli considerati, suggerendo che la sola esplicativa non riesce a spiegare una percentuale sufficiente di variabilità del modello.

Nella Figura 12 e nella Figura 13 vengono analizzati gli stessi modelli con l'ausilio di variabili di controllo che potrebbero spiegare meglio la relazione con gli indicatori di performance.

Figura 12

Risultati dei Modelli con Variabili di Controllo						
Dependent variable:						
	Redditività_delle_vendite_ROS			Redditività_del_capitale_proprio_ROE		
	(1)	(2)	(3)	(4)	(5)	(6)
CR4	-0.030 (0.047)			-0.235* (0.133)		
HHI		-0.0003 (0.0003)			-0.002* (0.001)	
Gini			-26.817*** (10.075)			-64.791** (28.810)
Debt_Equity_ratio	-0.007 (0.076)	-0.011 (0.074)	-0.012 (0.072)	0.059 (0.212)	0.009 (0.207)	-0.015 (0.205)
log_Totale_Attività	3.164** (1.414)	3.683** (1.512)	3.479** (1.380)	2.816 (3.963)	5.625 (4.241)	3.551 (3.946)
log_TOTALE_DEBITI	-2.610* (1.348)	-2.892** (1.370)	-2.485* (1.312)	0.500 (3.780)	-1.104 (3.842)	0.642 (3.750)
log_Capitale_sociale	-0.122 (0.700)	-0.177 (0.686)	-0.188 (0.668)	-2.857 (1.961)	-3.351* (1.925)	-3.462* (1.911)
Observations	135	135	135	135	135	135
R2	0.046	0.050	0.098	0.055	0.058	0.071
Adjusted R2	-0.112	-0.107	-0.051	-0.101	-0.097	-0.083
F Statistic (df = 5; 115)	1.107	1.218	2.503**	1.345	1.422	1.746
Note: *p<0.1; **p<0.05; ***p<0.01						

Figura 13

	Redditività_del_totale_attivo_ROA			EBITDA_Vendite		
	(7)	(8)	(9)	(10)	(11)	(12)
CR4	-0.016 (0.117)			0.252 (0.302)		
HHI		-0.001 (0.001)			-0.001 (0.002)	
Gini			-46.190* (25.274)			-163.514** (64.464)
Debt_Equity_ratio	0.335* (0.187)	0.340* (0.183)	0.339* (0.180)	0.625 (0.483)	0.745 (0.472)	0.754 (0.458)
log_Totale_Attività	4.064 (3.498)	4.949 (3.742)	4.613 (3.462)	4.153 (9.011)	6.220 (9.672)	6.151 (8.830)
log_TOTALE_DEBITI	1.670 (3.336)	1.226 (3.391)	1.919 (3.290)	3.960 (8.594)	3.230 (8.764)	5.145 (8.391)
log_Capitale_sociale	-1.789 (1.731)	-1.787 (1.699)	-1.806 (1.677)	-5.624 (4.459)	-4.822 (4.392)	-4.827 (4.276)
Observations	135	135	135	135	135	135
R2	0.112	0.115	0.137	0.051	0.048	0.095
Adjusted R2	-0.035	-0.031	-0.006	-0.106	-0.110	-0.054
F Statistic (df = 5; 115)	2.904**	2.998**	3.652***	1.224	1.149	2.425**
Note:	*p<0.1; **p<0.05; ***p<0.01					

L'indice HHI diventa significativo al 10% per spiegare il ROE, così come l'indice di Gini nel modello del ROA. Le variabili di controllo risultano utili a spiegare la variabilità della performance nei tre modelli in cui il ROS è regredito sugli indici di concentrazione, ma ad eccezione dell'indice di Gini essi non guadagnano significatività. Nello specifico, il logaritmo del totale dell'attività è significativo in tutti i casi al 5% e figura con coefficiente positivo: le aziende più grandi sembrano avere una maggiore redditività delle vendite. Anche il logaritmo del totale dei debiti è significativo in questi tre modelli ma il suo coefficiente è negativo, poiché un maggiore livello di indebitamento riduce la redditività.

Le variabili di controllo sembrano dunque spiegare meglio la variabilità degli indicatori di performance ma l' $R^2$  rimane ridotto, anche se leggermente maggiore.

Risulta allora utile considerare dei test diagnostici sui modelli più significativi per comprendere come migliorarli per spiegare meglio la relazione tra performance e concentrazione nel settore dei trasporti.

## 5.4 Test diagnostici

I test di diagnostica vengono svolti sui modelli che sembrano più significativi: ROS e ROE regrediti sull'indice di Gini con e senza variabili di controllo ed EBITDA/Vendite regredita con l'esplicativa Gini e le variabili di controllo. I p-value dei vari test sono riportati nella Tabella 2.

Tabella 2

**p-value dei test sui modelli con almeno un coefficiente significativo al 5%**

0.4235	0.0040	0.0000	0.2416	ROS CR4 ctrl
0.7949	0.0035	0.0000	0.2950	ROS HHI ctrl
0.8736	0.0005	0.0009	0.5814	ROS Gini ctrl
0.6495	0.0002	0.0000	0.2183	ROS Gini noctrl
0.0271	0.4761	0.0095	0.7540	ROE Gini ctrl
0.3699	0.6383	0.0000	0.8788	ROE Gini noctrl
0.2764	0.2159	0.0002	0.0001	EBITDA Gini ctrl
Hausman_Test	BP_Effects_Test	Wooldridge_Test	BP_Hetero_Test	

Gli esiti sono i seguenti:

- Per quanto riguarda il modello di ROS e indice di Gini senza variabili di controllo, i risultati sembrano suggerire che non vi siano differenze significative tra la scelta di effetti casuali o fissi, portando dunque a preferire un modello a effetti casuali. Un modello panel, comunque, non sembra necessario e un OLS potrebbe essere adeguato. Vi è inoltre presenza di autocorrelazione e omoschedasticità;
- Il modello del ROE regredito sull'indice di Gini senza variabili di controllo presenta residui omoschedastici e autocorrelati, mentre un modello panel sembra essere adeguato con effetti casuali preferibili a quelli fissi;
- Per quanto riguarda il modello che regredisce il ROS sull'indice di Gini e le variabili di controllo, si preferiscono gli effetti casuali ma potrebbero non essere significativi e i residui sono sempre autocorrelati e omoschedastici;



- Nel caso in cui nel modello di ROE e Gini si aggiungano le variabili di controllo, è raccomandato un modello panel ad effetti fissi e i residui rimangono autocorrelati e omoschedastici;
- Nel modello dell'EBITDA/Vendite considerato, i test portano a preferire un modello panel con effetti casuali, ma i residui sono eteroschedastici oltre che autocorrelati.

Altri due modelli, sebbene non aggiungano significatività agli indici di concentrazione  $CR_4$  e HHI, sembrano spiegare meglio il ROS grazie all'impiego delle variabili strumentali:

- Nel caso di modello di  $CR_4$  e variabili di controllo, gli effetti fissi non sono preferibili a quelli casuali, che comunque non risultano significativi. I residui sono omoschedastici e autocorrelati;
- Si conferma la preferenza per un modello a effetti casuali che però non sono significativi rispetto a un OLS nel caso la variabile esplicativa sia l'HHI e i residui sono ancora una volta omoschedastici e autocorrelati.

## 6. CONCLUSIONI

L'economia industriale studia il comportamento delle imprese in una data struttura di mercato, le determinanti del potere di mercato e le ragioni alla base dello scostamento delle condizioni di equilibrio da quelle di concorrenza perfetta.

Per effettuare analisi di questo tipo, è stato necessario innanzitutto definire la struttura di mercato, che comprende quindi anche l'esplorazione della sua concentrazione. Si tratta del primo passo necessario a uno studio di economia industriale quindi risulta estremamente cruciale.

La ricerca proposta in questo elaborato si è occupata di indagare proprio la concentrazione di mercato e una sua possibile relazione con le performance aziendali all'interno del mercato dei trasporti italiano (sezione H della nomenclatura NACE Rev. 2).

Il settore considerato è estremamente ampio; per rendere l'analisi più dettagliata, sono stati dunque estrapolati 15 sottosettori sulla base dei primi tre valori della classificazione NACE.

Sulla base dei risultati ottenuti si può concludere che nel settore considerato non sembra emergere un forte legame tra indici di concentrazione di mercato e indicatori di performance aziendali.

Dall'analisi della correlazione, è emerso un legame abbastanza forte e positivo solo tra l'indice  $CR_4$  e il logaritmo del totale dell'attività, del totale del debito e del capitale sociale delle aziende, variabili che risultano correlate anche con l'indice HHI. Queste tre variabili economiche colgono prevalentemente le dimensioni aziendali e sono meno indicative della vera e propria performance, catturata invece da indicatori quali ROA, ROE, ROS ed EBITDA/Vendite.

Si potrebbe quindi affermare che vi è principalmente una relazione tra concentrazione di mercato e dimensione aziendale, relazione che dai modelli esaminati risulta positiva. Per quando riguarda invece il legame tra  $CR_4$ , HHI e indice di Gini con gli indicatori di performance, sembra emergere una correlazione di entità molto bassa e di segno negativo, suggerendo che all'aumentare della concentrazione in un settore segue una diminuzione delle prestazioni aziendali.

Questa considerazione è stata anche confermata dall'analisi successiva, svolta mediante l'utilizzo di modelli di regressione lineare.

I modelli di regressione a effetti fissi non hanno evidenziato delle relazioni significative fra concentrazione e performance; gli unici coefficienti significativi riguardano l'indice di Gini con ROS, ROE, ROA e EBITDA/Vendite. In tutti e tre questi modelli è stato stimato un legame negativo, il quale indica che, se l'indice di Gini aumenta (maggior concentrazione), gli indicatori di performance si riducono.

Questi modelli però non hanno una buona bontà di adattamento ai dati empirici; per questo motivo sono stati proposti dei test diagnostici per cercare di capire quali modelli sono più appropriati per indagare la relazione tra le varie coppie di indici di concentrazione e performance. Questo ultimo passaggio può essere preso in considerazione come punto di partenza di studi futuri sul legame tra i singoli indicatori di concentrazione con i vari indici di performance.

## 7. BIBLIOGRAFIA E SITOGRAFIA

- Smirlock, M. (1985). Evidence on the (non) relationship between concentration and profitability in banking. *Journal of money, credit and Banking*, 17(1), 69-83.
- Weiss, L. W. (1974). Cowling's Market Structure and Corporate Behavior.
- Strickland, A. D., & Weiss, L. W. (1976). Advertising, concentration, and price-cost margins. *journal of political Economy*, 84(5), 1109-1121.
- Phillips, A. (1986). Market concentration and performance: A survey of the evidence. *Notre Dame L. Rev.*, 61, 1099.
- Kwoka, J. E. (1979). The effect of market share distribution on industry performance. *The Review of Economics and Statistics*, 101-109.
- Lee, S. (2014). The relationship between growth and profit: evidence from firm-level panel data. *Structural Change and Economic Dynamics*, 28, 1-11.
- Salinger, M. A. (1984). Tobin's q, unionization, and the concentration-profits relationship. *the Rand journal of Economics*, 15(2), 159-170.
- Černohorský, J., & Prokop, V. (2016). The relationship of concentration and profitability in banking markets. In *15th International Conference on Finance and Banking in Silesian University.[Conference proceedings] Karviná: OPF v Karviné* (pp. 40-49).
- Ornstein, S. I. (1972). Concentration and profits. *The Journal of Business*, 45(4), 519-541.
- Chok, N. S. (2010). *Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data* (Doctoral dissertation, University of Pittsburgh).
- Counsell, A., & Cribbie, R. A. (2015). Equivalence tests for comparing correlation and regression coefficients. *British Journal of Mathematical and Statistical Psychology*, 68(2), 292-309.
- Asuero, A. G., Sayago, A., & González, A. G. (2006). The correlation coefficient: An overview. *Critical reviews in analytical chemistry*, 36(1), 41-59.
- Curry, B., & George, K. D. (1983). Industrial concentration: a survey. *The Journal of Industrial Economics*, 31(3), 203-255.
- DiCiccio, C. J., & Romano, J. P. (2017). Robust permutation tests for correlation and regression coefficients. *Journal of the American Statistical Association*, 112(519), 1211-1220.
- Chatzipantsiou, C., Dimitriadis, M., Papadakis, M., & Tsagris, M. (2018). Extremely efficient permutation and bootstrap hypothesis tests using R. *arXiv preprint arXiv:1806.10947*.
- Sheytanova, T. (2015). The accuracy of the Hausman Test in panel data: A Monte Carlo study.

- Di Iorio, G., De Angelis, M., & Sofronic, B. (2021). *Sophia: ontologia iniziale del progetto*.
- Borgarello, M., & Fuso Nerini, F. (2011). Analisi dei consumi energetici dei settori industriali. *Ricerca sistema energetico*.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research synthesis methods*, 1(2), 97-111.