

Abstract

This study analyzes cross-country differences in aquaculture production using a panel dataset covering the period 2013–2022. Data from OECD, Eurostat, and FAO sources are combined to construct a country–year dataset including aquaculture output, GDP, employment, trade, patents, and environmental indicators. The main objective is to examine whether countries with higher GDP tend to exhibit higher aquaculture production, reflecting greater economic capacity to invest in aquaculture activities. Exploratory data analysis is used to identify patterns and relationships in the data prior to further statistical analysis.

1. Introduction

Aquaculture has become an increasingly important source of global food production as capture fisheries face biological and environmental constraints. Understanding the factors associated with aquaculture production is therefore important for both economic and policy analysis. This study focuses on the role of macroeconomic capacity, particularly GDP, in explaining cross-country differences in aquaculture output. Rather than making causal claims, the analysis aims to identify systematic relationships consistent with economic intuition using observational data.

2. Data and Analytical Framework

2.1 Data Sources

Data were collected from publicly available international databases. Aquaculture production data were obtained from the OECD Data Explorer and measured in live weight (tonnes). Macroeconomic indicators such as GDP and population were sourced from OECD datasets. Fisheries trade data (imports and exports) and fleet and catch statistics were obtained from Eurostat. Information on fishing- and aquaculture-related patents was used as a proxy for technological activity, while FAO fisheries membership and coastal access were included as binary indicators.

2.2 Data Preparation and Panel Construction

All datasets were harmonized at the country–year level for the period 2013–2022. Country identifiers were standardized using ISO3 country codes, and only countries with sufficient data coverage were retained. Extensive data cleaning was performed to handle missing values, inconsistent formats, and reporting flags. Variables with unclear source traceability, such as

industry productivity, were excluded to ensure transparency and reproducibility. The final dataset represents a well-documented panel suitable for descriptive and regression-based analysis.

2.3 Tools Used

The analysis was conducted using Python. Data cleaning and merging were performed with the pandas library, while numerical operations relied on NumPy. Visualization and exploratory analysis were supported by standard plotting libraries. The workflow was designed to ensure transparency and reproducibility.

3. Descriptive Exploration of the Data

Exploratory data analysis was conducted to examine distributions and relationships among key variables. Initial visualizations suggest a positive association between GDP and total aquaculture production, consistent with the proposed hypothesis. Similar patterns are observed for aquaculture employment and fisheries exports. These descriptive findings provide motivation for further statistical analysis while remaining exploratory in nature.

3.1 Data Visualization

Correlation heatmaps were used to examine relationships among key fisheries, aquaculture, economic, and innovation variables. The heatmap of selected fisheries-related features indicates that aquaculture employment is moderately associated with aquaculture production, while variables related to capture fisheries such as fleet size and fish catches show weaker relationships with aquaculture output.

A broader correlation heatmap including all numeric variables highlights strong associations between macroeconomic indicators such as GDP and total patent counts, as well as between GDP and fisheries trade variables. In contrast, fishing-specific patent activity appears less strongly linked to overall economic size.

GDP trends across countries over time were visualized using a multi-line time-series plot. The visualization shows substantial differences in economic scale and growth patterns across countries, with a small number of countries consistently dominating global GDP levels.

Finally, a bar chart comparing total patents and fishing-related patents for the top patent-producing countries reveals that fisheries-related innovation constitutes a relatively small share of overall patent activity, even among technologically advanced economies. A logarithmic scale was used to enable clearer comparison across countries with large differences in patent volumes.

Machine Learning Analysis

To complement the hypothesis testing, supervised machine learning was applied to evaluate how well macroeconomic, sectoral, and environmental variables jointly explain aquaculture production across countries. The aim of this analysis was predictive and exploratory rather than causal, focusing on identifying key drivers of production levels.

Data Preparation and Missing Values

The dataset contained substantial missing values in several variables. Initial row-wise deletion of missing observations resulted in an extremely small sample size, rendering the analysis unreliable. To address this issue, a column-wise missing data strategy was adopted. Variables with excessively high proportions of missing values were excluded, while variables with relatively low missingness were retained. This approach preserved a sufficient number of observations while ensuring model stability. After this process, the final machine learning dataset consisted of 360 observations.

Aquaculture production, measured in total tonnes, was used as the dependent variable. Due to strong right-skewness, the variable was log-transformed to stabilize variance and reduce the influence of extreme values.

Model Specifications

Two linear regression models were estimated. The first, a full model, included GDP, population, fish catches, fisheries exports, fisheries imports, and exposure to extreme temperatures. This model achieved an R^2 of approximately 0.54, indicating that over half of the variation in aquaculture production is explained by the included variables. For cross-country economic data, this level of explanatory power is considered strong.

To assess robustness and address multicollinearity, an alternative specification excluding population was also estimated. Population was removed due to its very high correlation with GDP. The alternative model achieved a lower but still meaningful R^2 of approximately 0.40, confirming that population plays a substantial role in explaining production levels.

Interpretation of Results

In the full model, population emerged as the strongest predictor, reflecting a clear scale effect: countries with larger populations tend to produce more aquaculture output. GDP exhibited a negative coefficient in this specification; however, this result is driven by multicollinearity with population and should be interpreted conditionally rather than independently. It does not imply

that higher GDP reduces aquaculture production, but rather that GDP adds limited explanatory power once demographic and sectoral factors are controlled for.

In the alternative model without population, GDP became positive but economically small, while fisheries exports emerged as the most influential variable. This suggests that aquaculture production is more closely linked to fisheries-sector orientation and trade activity than to income level alone. Environmental exposure to extreme temperatures showed a moderate association, likely capturing regional patterns rather than direct causal effects.

Evaluation

Model performance was evaluated using RMSE and R^2 , which are appropriate metrics for continuous outcomes.

Conclusion

This study examined the relationship between macroeconomic conditions and aquaculture production across countries by combining hypothesis testing and machine learning approaches. The central hypothesis proposed that countries with higher GDP tend to have higher aquaculture production due to greater investment capacity and economic resources.

The empirical analysis shows that this relationship is more nuanced than a simple positive association. While GDP displays a positive relationship with aquaculture production in simpler model specifications, its independent effect weakens once demographic and sector-specific variables are introduced. In particular, population emerges as the most influential factor, indicating that aquaculture production is strongly driven by scale effects. Countries with larger populations tend to exhibit substantially higher aquaculture output, regardless of income level.

The machine learning results reinforce this conclusion. The full linear regression model explains approximately 54% of the variation in aquaculture production, highlighting the importance of demographic size, fisheries trade, and sectoral characteristics. An alternative model excluding population shows reduced explanatory power, confirming the dominant role of scale. GDP, while relevant, appears to operate conditionally rather than as a primary driver.

Overall, the findings suggest that aquaculture production is shaped more by population size and fisheries-sector structure than by economic capacity alone. Rather than rejecting the initial hypothesis, the results provide partial support by demonstrating that GDP matters primarily through its interaction with demographic and industry-related factors.

Limitations and Future Work

This study is subject to several limitations. First, the analysis relies on country-level aggregate data, which may mask important regional and local differences in aquaculture practices.

Within-country variation in infrastructure, regulation, and environmental conditions cannot be captured in this framework.

Second, missing data posed a significant challenge. Although a column-wise approach to handling missing values preserved sample size and ensured model stability, excluding variables with high missingness may have omitted potentially relevant explanatory factors. As a result, the models may not fully capture technological or institutional dimensions of aquaculture development.

Third, the machine learning analysis focuses on linear relationships. While linear models offer interpretability and are appropriate for hypothesis-driven analysis, they may fail to capture complex non-linear dynamics inherent in environmental and economic systems.

Future research could address these limitations in several ways. Incorporating regional or subnational data would allow for more granular analysis of aquaculture production. Improved data availability on technological adoption, environmental regulation, and sustainability practices could enhance explanatory power. Additionally, extending the analysis to non-linear or panel-based modeling approaches could provide deeper insights into dynamic and long-term relationships.

Despite these limitations, the study provides a robust and methodologically sound examination of the drivers of aquaculture production and highlights the importance of scale and sectoral structure in shaping outcomes.

Bibliography

Artificial intelligence tools (ChatGPT by OpenAI) were used to assist with coding and to provide guidance on structuring the analysis and writing the report. The AI tool served as a supportive resource

[https://data-explorer.oecd.org/vis?fs\[0\]=Topic%2C1%7CAgriculture%20and%20fisheries%23AGR%23%7CFisheries%20and%20aquaculture%23AGR_FSA%23&pg=0&fc=Topic&bp=true&snb=6&vw=tb&df\[ds\]=dsDisseminateFinalDMZ&df\[jd\]=DSD_FISH_PROD%40DF_FISH_AQUA&df\[ag\]=OECD.TAD.ARP&df\[vs\]=1.0&dq=.A.._T.T&pd=2013%2C&to\[TIME_PERIOD\]=false](https://data-explorer.oecd.org/vis?fs[0]=Topic%2C1%7CAgriculture%20and%20fisheries%23AGR%23%7CFisheries%20and%20aquaculture%23AGR_FSA%23&pg=0&fc=Topic&bp=true&snb=6&vw=tb&df[ds]=dsDisseminateFinalDMZ&df[jd]=DSD_FISH_PROD%40DF_FISH_AQUA&df[ag]=OECD.TAD.ARP&df[vs]=1.0&dq=.A.._T.T&pd=2013%2C&to[TIME_PERIOD]=false)

[https://data-explorer.oecd.org/vis?lc=en&fs\[0\]=Topic%2C1%7CEnvironment%20and%20climate%20change%23ENV%23%7CTechnology%20and%20innovation%23ENV_TEC%23&pg=0&fc=Topic&bp=true&snb=5&vw=tb&df\[ds\]=dsDisseminateFinalDMZ&df\[jd\]=DSD_PAT_DEV%40DF_PAT_DEV&df\[ag\]=OECD.ENV.EPI&df\[vs\]=1.0&dq=.A.GOODS_AGR_FISH%2BTOT.PATN.TWO.&pd=2013%2C&to\[TIME_PERIOD\]=false](https://data-explorer.oecd.org/vis?lc=en&fs[0]=Topic%2C1%7CEnvironment%20and%20climate%20change%23ENV%23%7CTechnology%20and%20innovation%23ENV_TEC%23&pg=0&fc=Topic&bp=true&snb=5&vw=tb&df[ds]=dsDisseminateFinalDMZ&df[jd]=DSD_PAT_DEV%40DF_PAT_DEV&df[ag]=OECD.ENV.EPI&df[vs]=1.0&dq=.A.GOODS_AGR_FISH%2BTOT.PATN.TWO.&pd=2013%2C&to[TIME_PERIOD]=false)

[https://data-explorer.oecd.org/vis?tm=gdp&pg=0&snb=263&vw=tb&df\[ds\]=dsDisseminateFinalDMZ&df\[id\]=DSD_NAMAIN10%40DF_TABLE1_EXPENDITURE&df\[ag\]=OECD.SDD.NAD&df\[vs\]=2.0&dq=A.AUT%2BBEL%2BCAN%2BCHL%2BCOL%2BCRI%2BCZE%2BDNK%2BEST%2BFIN%2BFRA%2BDEU%2BGRC%2BHUN%2BISL%2BIRL%2BISR%2BITA%2BJPN%2BKOR%2BLVA%2BLTU%2BLUX%2BMEX%2BNLD%2BNZL%2BNOR%2BPOL%2BPRT%2BSVK%2BSVN%2BESP%2BSWE%2BCHE%2BTUR%2BGBR%2BUSA%2BEA20%2BEU27_2020%2BALB%2BARG%2BBRA%2BBGR%2BCPV%2BCMR%2BCHN%2BHRV%2BCYP%2BGEO%2BHKG%2BIND%2BIDN%2BKAZ%2BMDG%2BMLT%2BMAR%2BMKD%2BROU%2BRUS%2BSAU%2BSEN%2BSRB%2BSGP%2BZAF%2BZMB%2BAUS...B1GQ...USD_EXC.V.&pd=2013%2C&to\[TIME_PERIOD\]=false](https://data-explorer.oecd.org/vis?tm=gdp&pg=0&snb=263&vw=tb&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_NAMAIN10%40DF_TABLE1_EXPENDITURE&df[ag]=OECD.SDD.NAD&df[vs]=2.0&dq=A.AUT%2BBEL%2BCAN%2BCHL%2BCOL%2BCRI%2BCZE%2BDNK%2BEST%2BFIN%2BFRA%2BDEU%2BGRC%2BHUN%2BISL%2BIRL%2BISR%2BITA%2BJPN%2BKOR%2BLVA%2BLTU%2BLUX%2BMEX%2BNLD%2BNZL%2BNOR%2BPOL%2BPRT%2BSVK%2BSVN%2BESP%2BSWE%2BCHE%2BTUR%2BGBR%2BUSA%2BEA20%2BEU27_2020%2BALB%2BARG%2BBRA%2BBGR%2BCPV%2BCMR%2BCHN%2BHRV%2BCYP%2BGEO%2BHKG%2BIND%2BIDN%2BKAZ%2BMDG%2BMLT%2BMAR%2BMKD%2BROU%2BRUS%2BSAU%2BSEN%2BSRB%2BSGP%2BZAF%2BZMB%2BAUS...B1GQ...USD_EXC.V.&pd=2013%2C&to[TIME_PERIOD]=false)

[https://data-explorer.oecd.org/vis?tm=population&pg=0&snb=302&vw=tb&df\[ds\]=dsDisseminateFinalDMZ&df\[id\]=DSD_POPULATION%40DF_POP_HIST&df\[ag\]=OECD.ELS.SAE&df\[vs\]=1.0&dq=AUT%2BBEL%2BCAN%2BCHL%2BCOL%2BCRI%2BCZE%2BDNK%2BEST%2BFIN%2BFRA%2BDEU%2BGRC%2BHUN%2BISL%2BIRL%2BISR%2BITA%2BJPN%2BKOR%2BLVA%2BLTU%2BLUX%2BMEX%2BNLD%2BNZL%2BNOR%2BPOL%2BPRT%2BSVK%2BSVN%2BESP%2BSWE%2BCHE%2BTUR%2BGBR%2BUSA%2BG20%2BEU27%2BOECD%2BARG%2BBRA%2BBGR%2BCHN%2BHRV%2BCYP%2BIND%2BIDN%2BMLT%2BROU%2BRUS%2BSAU%2BSGP%2BZAF%2BW%2BAUS.POP.PS._T._T.&pd=2013%2C2024&to\[TIME_PERIOD\]=false](https://data-explorer.oecd.org/vis?tm=population&pg=0&snb=302&vw=tb&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_POPULATION%40DF_POP_HIST&df[ag]=OECD.ELS.SAE&df[vs]=1.0&dq=AUT%2BBEL%2BCAN%2BCHL%2BCOL%2BCRI%2BCZE%2BDNK%2BEST%2BFIN%2BFRA%2BDEU%2BGRC%2BHUN%2BISL%2BIRL%2BISR%2BITA%2BJPN%2BKOR%2BLVA%2BLTU%2BLUX%2BMEX%2BNLD%2BNZL%2BNOR%2BPOL%2BPRT%2BSVK%2BSVN%2BESP%2BSWE%2BCHE%2BTUR%2BGBR%2BUSA%2BG20%2BEU27%2BOECD%2BARG%2BBRA%2BBGR%2BCHN%2BHRV%2BCYP%2BIND%2BIDN%2BMLT%2BROU%2BRUS%2BSAU%2BSGP%2BZAF%2BW%2BAUS.POP.PS._T._T.&pd=2013%2C2024&to[TIME_PERIOD]=false)

[https://data-explorer.oecd.org/vis?pg=0&bp=true&snb=23&tm=temperature&vw=tb&df\[ds\]=dsDisseminateFinalDMZ&df\[id\]=DSD_ECH%40EXT_TEMP_H&df\[ag\]=OECD.ENV.EPI&df\[vs\]=2.0&dq=AFG%2BBFA%2BBDI%2BCAF%2BTCD%2BSWZ%2BETH%2BGUF%2BATF%2BGLP%2BHMD%2BVAT%2BKAZ%2BXKV%2BKGY%2BLAO%2BLSO%2BLIE%2BMWI%2BMLI%2BMTQ%2BMDA%2BMNG%2BNPL%2BNER%2BNFK%2BMKD%2BPRY%2BREU%2BRWA%2BSMR%2BSRB%2BSGS%2BSSD%2BTJK%2BTKM%2BUGA%2BUMI%2BUZB%2BZMB%2BZWE%2BIOT%2BBVT%2BBWA%2BBOL%2BBTN%2BBLR%2BAZE%2BARM%2BAND%2BCHE%2BSVK%2BLUX%2BHUN%2BCZE%2BAUT%2BALB%2BDZA%2BASM%2BAGO%2BAIA%2BATG%2BARG%2BABW%2BBHS%2BBHR%2BGGY%2BBGD%2BBRB%2BBLZ%2BBEN%2BBMU%2BBIH%2BBRA%2BVGB%2BBRN%2BBGR%2BCPV%2BKHM%2BCMR%2BCYM%2BCHN%2BCXR%2BCOM%2BCOG%2BCOK%2BCIV%2BHRV%2BCUB%2BCYP%2BPRK%2BCOD%2BDJI%2BDMA%2BDOM%2BECU%2BEGY%2BSLV%2BGNQ%2BERI%2BFRO%2BFJI%2BPYF%2BGAB%2BGMB%2BGEO%2BGHA%2BGRL%2BGRD%2BGUM%2BGTM%2BGIN%2BGNB%2BGUY%2BHTI%2BHND%2BHKG%2BIND%2BIDN%2BIRN%2BIRQ%2BIMN%2BJAM%2BJEY%2BJOR%2BKEN%2BKIR%2BKWT%2BLBN%2BLBR%2BLBY%2BMAC%2BMDG%2BMYS%2BMDV%2BMLT%2BMHL%2BMRT%2BMUS%2BMYT%2BFSM%2BMCO%2BMNE%2BMSR%2BMAR%2BMOZ%2BMMR%2BNAM%2BNRU%2BNCL%2BNIC%2BNGA%2BNIU%2BMNP%2BOMN%2BPAK%2BPLW%2BPSE%2BPAN%2BPNG%2BPER%2BPHL%2BPCN%2BPRI%2BQAT%2BROU%2BRUS%2BSHN%2BKNA%2BLCA%2BSPM%2BVCT%2BWSM%2BSTP%2BSAU%2BSEN%2BSYC%2BSLE%2BSGP%2BSLB%2BSOM%2BZAF%2B](https://data-explorer.oecd.org/vis?pg=0&bp=true&snb=23&tm=temperature&vw=tb&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_ECH%40EXT_TEMP_H&df[ag]=OECD.ENV.EPI&df[vs]=2.0&dq=AFG%2BBFA%2BBDI%2BCAF%2BTCD%2BSWZ%2BETH%2BGUF%2BATF%2BGLP%2BHMD%2BVAT%2BKAZ%2BXKV%2BKGY%2BLAO%2BLSO%2BLIE%2BMWI%2BMLI%2BMTQ%2BMDA%2BMNG%2BNPL%2BNER%2BNFK%2BMKD%2BPRY%2BREU%2BRWA%2BSMR%2BSRB%2BSGS%2BSSD%2BTJK%2BTKM%2BUGA%2BUMI%2BUZB%2BZMB%2BZWE%2BIOT%2BBVT%2BBWA%2BBOL%2BBTN%2BBLR%2BAZE%2BARM%2BAND%2BCHE%2BSVK%2BLUX%2BHUN%2BCZE%2BAUT%2BALB%2BDZA%2BASM%2BAGO%2BAIA%2BATG%2BARG%2BABW%2BBHS%2BBHR%2BGGY%2BBGD%2BBRB%2BBLZ%2BBEN%2BBMU%2BBIH%2BBRA%2BVGB%2BBRN%2BBGR%2BCPV%2BKHM%2BCMR%2BCYM%2BCHN%2BCXR%2BCOM%2BCOG%2BCOK%2BCIV%2BHRV%2BCUB%2BCYP%2BPRK%2BCOD%2BDJI%2BDMA%2BDOM%2BECU%2BEGY%2BSLV%2BGNQ%2BERI%2BFRO%2BFJI%2BPYF%2BGAB%2BGMB%2BGEO%2BGHA%2BGRL%2BGRD%2BGUM%2BGTM%2BGIN%2BGNB%2BGUY%2BHTI%2BHND%2BHKG%2BIND%2BIDN%2BIRN%2BIRQ%2BIMN%2BJAM%2BJEY%2BJOR%2BKEN%2BKIR%2BKWT%2BLBN%2BLBR%2BLBY%2BMAC%2BMDG%2BMYS%2BMDV%2BMLT%2BMHL%2BMRT%2BMUS%2BMYT%2BFSM%2BMCO%2BMNE%2BMSR%2BMAR%2BMOZ%2BMMR%2BNAM%2BNRU%2BNCL%2BNIC%2BNGA%2BNIU%2BMNP%2BOMN%2BPAK%2BPLW%2BPSE%2BPAN%2BPNG%2BPER%2BPHL%2BPCN%2BPRI%2BQAT%2BROU%2BRUS%2BSHN%2BKNA%2BLCA%2BSPM%2BVCT%2BWSM%2BSTP%2BSAU%2BSEN%2BSYC%2BSLE%2BSGP%2BSLB%2BSOM%2BZAF%2B)

LKA%2BSDN%2BSUR%2BSJM%2BSYR%2BTWN%2BTZA%2BTHA%2BTLS%2BTGO%2BT
KL%2BTON%2BTTO%2BTUN%2BTCA%2BTUV%2BUKR%2BARE%2BVIR%2BURY%2BVUT
%2BVEN%2BVNM%2BWLF%2BESH%2BYEM%2BAU1%2BAU2%2BAU3%2BAU4%2BAU5%
2BAU6%2BAU7%2BAU8%2BAUS%2BBEL%2BCAN%2BCHL%2BCOL%2BCRI%2BDNK%2B
EST%2BFIN%2BFRA%2BDEU%2BGRC%2BISL%2BIRL%2BISR%2BITA%2BJPN%2BKOR%
2BLVA%2BLTU%2BMEX%2BNLD%2BNZL%2BNOR%2BPOL%2BPRT%2BSVN%2BESP%2B
SWE%2BTUR%2BGBR%2BUSA%2BOECD%2BOECD.A.HD_POP_EXP..W_LT_2.....&p
d=2013%2C&to[TIME_PERIOD]=false&isAvailabilityDisabled=false