Birmingham City
Business School

BIRMINGHAM CITY
University

CAFÉ WORKING PAPER NO. 30

Centre for Accountancy, Finance, and Economics (CAFÉ)

Determinants of Export Diversification in Resource-Dependent Economies: The Role of Product Relatedness and Macroeconomic Conditions

Beatriz Calzada Olvera and Danilo Spinola

September 2024

Determinants of Export Diversification in Resource-Dependent Economies: The Role of Product Relatedness and Macroeconomic Conditions

Beatriz Calzada Olvera

Erasmus University Rotterdam – Institute for Housing and Urban Development & UNU-MERIT.

E-mail: calzadaolvera@ihs.nl; calzada@merit.unu.edu

Danilo Spinola

Birmingham City University. College of Accountancy, Finance and Economics.

E-mail: danilo.spinola@bcu.ac.uk

Abstract

Export diversification is crucial for economic development, yet many resource-rich countries have struggled to achieve significant progress in diversifying its economic structure. While the lack of capabilities is often highlighted as a primary barrier to diversification, the literature frequently underestimates the significant impact of macroeconomic conditions on diversification potential. This study seeks to bridge the gap between the capabilities literature and macroeconomic factors, particularly in the context of economies heavily dependent on extractive industries. In order to address our question, we initially introduce a novel measure of product relatedness, expanding on the framework developed by Nomaler and Verspagen (2022), and econometrically estimate its relationship with key macroeconomic variables such as international prices, exchange rates, energy and mineral dependency, and GDP per capita. The analysis spans over 5,000 products across multiple countries from 1995 to 2019, with the objective of determining the relative significance of these factors in predicting diversification patterns and assessing how macroeconomic conditions either facilitate or impede diversification, particularly in non-extractive sectors. Product relatedness predicts diversification, especially in extractive industries where path dependence is highly pronounced. However, macroeconomic factors exert a major influence on diversification outcomes. These macroeconomic variables can either constrain or enable diversification, shaping the pathways through which industries evolve and expand their portfolios.

Keywords: Export Diversification, Product Relatedness, Macroeconomic Factors, Extractive Sectors

JEL Codes: O11, F14, L78

1. Introduction

Economic diversification is widely recognised as essential for sustainable development, particularly in countries rich in extractive resources. The ability to produce a varied range of technologically sophisticated goods and services is crucial for reducing dependency on volatile commodity markets and ensuring long-term economic stability (Hausmann et al., 2005; Perez, 1999). While short-term growth can be achieved through reliance on extractive resources, sustainable, inclusive growth necessitates the development of a diversified and technologically advanced economy (Diao et al., 2019).

However, many resource-rich countries have struggled to diversify their economies. Between 1980 and 2010, export concentration intensified in most oil and mineral-producing nations (Ross, 2019). The decline in commodity prices during the mid-2010s rekindled interest in the relationship between extractive industries and diversification, highlighting the need for a deeper understanding of the factors that either promote or inhibit diversification efforts (Erten & Ocampo, 2021).

Existing literature offers varied perspectives on the determinants of diversification. The evolutionary economic geography literature, notably the work of Hausmann and Hidalgo (2014), focuses on leveraging existing capabilities to achieve diversification. This approach, which developed into a whole literature on the concept of relatedness, suggests that countries can more effectively diversify by developing products closely related to those they already produce, thereby capitalizing on established resources and knowledge (Boschma & Capone, 2015). In contrast, However, critics argue that this approach often overlooks other critical factors, such as institutional quality, infrastructure, and the combination of production factors, which also influence relatedness among industries (Guo & He, 2017).

Another perspective in the literature explores the critical role of macroeconomic constraints in shaping economic diversification. Recent studies argue that macroeconomic and trade-related factors—such as fiscal space, interest rate management, and exchange rate dynamics—are significant barriers to diversification (Botta et al., 2024; Porcile et al., 2022; Bresser-Pereira, 2020). Within this framework, neo-structuralist approaches highlight the importance of real exchange rate appreciation, the structure and stability of financial systems, and fluctuations in global commodity prices as pivotal elements influencing a country's potential to diversify its economy (Cimoli et al., 2016; Porcile et al., 2022; Guzman et al., 2018).

Nevertheless, a significant research gap remains in the empirical exploration of how product relatedness and macroeconomic factors interact to influence diversification. While both areas have been studied independently, there is limited evidence on how these two approaches can be integrated to provide a more comprehensive understanding of diversification dynamics. This paper aims to fill the research gap by investigating the determinants of product diversification, specifically focusing on the role of related variety, macroeconomic factors, and commodity dependence. We seek to answer the following research questions:

- How does related variety influence the development of comparative advantage in non-extractive versus extractive products?
- What impact do macroeconomic variables, such as real exchange rates and commodity prices, have on the likelihood of diversifying into non-extractive products?
- To what extent do macroeconomic factors mediate the relationship between related variety and diversification in non-extractive products?

The paper is structured as follows: Section 2 reviews the theoretical and empirical literature on export diversification, particularly in the context of natural resources. Section 3 outlines the methodology and data used in the analysis. Section 4 presents the empirical results, and Section 5 concludes with a discussion of the findings and potential avenues for future research.

2. Theoretical and Empirical Background

Recent empirical research in economic development (Hausmann et al., 2005; McMillan et al., 2014) has re-emphasized key insights from classical structuralist thought, highlighting the critical role of export composition and diversity in driving economic progress. Particularly for resource-rich nations specialising in minerals and energy, export diversification is seen as a vital strategy to mitigate the risks of price volatility, foster sustainable long-term growth, expand employment opportunities beyond the resource sector, and prepare for the eventual depletion of natural resources (Ross, 2019). Rising global efforts to reduce greenhouse gas emissions by consuming fewer fossil fuels make diversification among oil and gas exporters even more pressing.

Concerning the general economic benefits of diversification, several papers have identified a positive empirical association between export diversification and economic growth (Al-Marhub, 2000; Klinger and Lederman, 2006; Hesse, 2008) The latter two find that the relationship between export diversification and per capita income growth follows an inverted-U function, implying that countries get higher returns from diversifying their exports at lower levels of economic development than at very high ones.

To explain the positive relationship between diversification and growth several scholars have provided theoretical underpinnings – typically linking diversification to innovative activity. From an evolutionary economic perspective, innovation primarily involves the recombination of existing ideas into new configurations (Nelson & Winter, 1982). Additionally, innovation relies on a certain level of tacit, context-specific knowledge, which is often difficult to transfer across borders (Maskell & Malmberg, 1999). As a result, a country's productive structure and technological trajectory are highly path-dependent: what a country produces today significantly influences its future production capabilities (Dosi et al., 1990; Nelson & Winter, 1982).

The Evolutionary economic geography literature builds upon the latter idea to explain diversification patterns: a country will produce (and export) new products largely like those it already produces. This is because producing such new products requires productive capabilities, i.e., resources, knowledge, and capacities similar to those that the country already possesses (Hidalgo et al., 2007). In this view, if we consider two products, the possibility of becoming specialized in one (given specialization in the other) depends on whether they require the same capabilities – in other words, it depends on whether those two products relate (or not) in terms of productive capabilities. Studies in this strand have established that product relatedness[1] is a determinant of diversification – either at national or regional levels (Boschma et al., 2012; Hausmann & Klinger, 2007; Neffke et al., 2011). They show, in other words, that diversification patterns are highly path-dependent. Nonetheless, as pointed out in Boschma and Capone (2015) these studies do not explain differences in the diversification patterns across countries. Indeed, product-relatedness measures employed in such studies (i.e., Hausmann & Klinger, 2007) rely on export co-occurrence to proxy for similar productive capabilities, but they do not explain *why* those goods are exported in some countries and not in others (Content & Frenken, 2016).

To learn more about the determinants of the direction and intensity of the diversification processes, more recent empirical frameworks have then incorporated the role of institutions and governance (e.g. Boschma & Capone, 2015; He & Zhu, 2018), as well as global linkages captured by imports, FDI, and/or trade liberalization (Alonso & Martín, 2019; He et al., 2018) to shed further light on explaining differences. Most of these studies, however, have focused on within-country determinants.

A knowledge gap remains concerning the factors that play a role in the emergence and development of productive capabilities and, more specifically, those that enable entrepreneurs to engage in innovation activities, ultimately leading to diversification. According to Lall (1992), a country's technological capabilities are determined by the interplay of general capabilities (e.g., human capital); institutions, and incentives stemming from competition, factor markets, and naturally, macroeconomic factors, such as price changes, exchange rates, credit and foreign exchange availability, political stability or exogenous shocks (e.g., terms of trade). The following paragraphs focus on discussing some of the macroeconomic (and other country) characteristics that have been empirically tested in previous studies.

As pointed out by several scholars (Agosin et al., 2012; Alsharif et al., 2017; Ross, 2019; Wiig & Kolstad, 2012), even though diversification has been prescribed as essential in boosting economic development, the strategies

[1] This namely refers to the product relatedness measures developed in Hausmann and Klinger (2007) and Hidalgo et al. (2007) which have been widely employed in that type of empirical analysis. Yet there are other measures capturing how related productive capabilities of different products are; for instance, Franken et al. (2007) who look at the hierarchical classification of products by the SIC scheme.

and pathways for achieving it remain complex and open to various interpretations. Scholarly empirical works on the determinants of diversification have identified some inhibiting factors, such as natural resource abundance; but the role that key macroeconomic factors play, such as investment, interest rate and real exchange rate, is still inconclusive.

Esanov (2012), using a panel random-effects framework covering the 1980-2006 period, finds that export concentration is positively related to the share of natural resources in total exports; contrariwise, the study suggests a negative correlation of concentration with investment and trade freedom but no correlation with trade openness, inflation, FDI, or quality of institutions. Ahmadov (2014) using an IV setup which looks at the 1970-2010 period, further confirms that diversification is negatively associated with countries rich in resources but that this result applies only to countries that are rich in oil, located in Africa or the Middle East. No significant effects are found for human capital, trade openness, and quality of government. Along the same lines, Bahar and Santos (2016), using a variety of non-resource export concentration indices for the period 1985-2010, find strong evidence that higher shares of natural resources are associated with lower non-resource export diversification. Finally, Alsharif et al. (2017) find that oil exports are negatively associated with diversification (in this case, measured by non-oil rents). These studies thus provide empirical evidence that the more a country depends on commodity resources, the less likely it is to diversify its basket of exports.

Empirical studies have explored the causal link between the real exchange rate and diversification, particularly focusing on currency exchange misalignments like overvaluation, which is central to the Dutch disease argument (Corden & Neary, 1982). Higher commodity prices often lead to increased exports in booming sectors, resulting in substantial foreign exchange inflows and real currency appreciation. This, in turn, reduces the competitiveness of other tradable goods, driving further specialisation in the booming sector. Rodrik (2008) argues that currency undervaluation can promote diversification in weak institutional contexts by acting as a production subsidy and consumption tax on tradables. However, empirical evidence on the relationship between real exchange rates and export diversification remains mixed. Sekkat (2016) found that while currency undervaluation positively affects the share of manufactures in total exports, misalignment does not significantly impact export concentration, even in countries with low institutional quality. Tran et al. (2017) identified real exchange rates as a determinant of export diversification in only three developing countries, with a broader sample showing bi-directional causality. Agosin et al. (2012) observed that while exchange rate overvaluation does not significantly affect diversification, increasing terms of trade negatively impact it by reallocating factors to the booming commodity sector, reducing inputs for new product exports. This suggests that commodity price increases may influence export concentration through factor reallocation rather than solely through real exchange rate movements. This also resonates with relatively recent commodity price trends. As pointed out in UNCTAD (2019), rising commodity prices between 1998 to 2017 contributed to changes in the export composition of commodity exporters – changes which typically consisted of further export of concentration in oil and, especially, mineral exports[2] (UNCTAD, 2019).

Considering the discussion above, the current analysis combines empirical literature which looks at diversification at the product level and macroeconomic variables, namely real exchange rate, prices, and export dependence – given their relevance for understanding the dynamics of extractive and non-extractive exports. Looking at product level diversification in the empirical framework instead of export concentration – which is a measure that can be easily contaminated by price fluctuations (Alsharif et al., 2017)[3] - and using an alternative measure for relatedness, this study sheds further light on how path dependence predicts diversification in non-extractive and extractive goods.

A final consideration is that diversification in extractive commodities has received little empirical attention in recent years. Yet, not a lot is known about the determinants of this process: certainly, being able to diversify into extractive commodities is to a large extent 'God-given', but modern extractive resource industries often demand non-trivial technological, economic, political, and social processes (Ville & Wicken, 2012). Therefore, understanding how path dependency and macroeconomic factors play out for extractive products vis-à-vis non-extractives may also contribute to understanding the overall dynamics of diversification processes.

3. Methodology

3.1 Related Variety Calculation

To build our explanatory variable, we build on a probability-based relatedness measure for related variety to account for diversification potential, as developed in Nomaler and Verspagen (2022). We adopt and develop this alternative measure to address the criticism raised by Nomaler and Verspagen (2024), who argue that density fails to adequately capture the concept of relatedness. Diversification in this context is defined as the increment in the number of products that a country exports with revealed comparative advantage (RCA)[4]. Akin to other commonly applied product relatedness measures, the measure we employ builds upon the idea that a

[2] Commodity-dependent countries increased from 92 in 1998–2002 to 102 in 2013–2017. Yet, countries dependent upon agricultural exports went from 50 to 37 between these two periods. In contrast, mineral-dependent countries increased from 14 to 33, and the number of energy-dependent countries rose from 28 to 32. According to the classification of UNCTAD (2019), a country is commodity-dependent when more than 60% of its total exports are comprised of commodities.

[3] Measuring diversification, can be problematic when looking at commodities. As pointed out in Alsharif et al. (2017), export concentration (i.e., commodity exports as a share of total exports) in the presence of a negative price shock could reflect a "pseudo diversification" process rather than genuine changes in the export composition.

[4] The method presented is an adapted version to method employed in the development of the Upgrading Triangle presented in Annex 7.2 of the Greater Mekong Subregion 2030 and Beyond Report (ADB, 2021).

country's ability to develop new products in the future is – at least in part – determined by its present specialisation structure.

First, $\boldsymbol{X}$ represents a binary matrix of RCA[5] with dimensions $m \times n$, where m corresponds to the number of products and n is the number of countries. A typical element in $\boldsymbol{X}$, represented by x_{ij}, takes a bivariate value, following the definition of RCA originally proposed by Balassa (1965):

$$x_{ij} = \begin{cases} 1 \; if \; RCA \geq 1 \\ 0 \; otherwise \end{cases}$$

Further, a conditional probability (product-by-product) matrix, $\boldsymbol{G}$, is defined in the following manner:

$$\boldsymbol{G} = (\boldsymbol{XX'})/s$$

where $\boldsymbol{X'}$ represents a transposed matrix and s is the vector containing the row-sum of $\boldsymbol{X}$ (i.e., the number of total exported products with comparative advantage by a given country)[6]. $\boldsymbol{G}$ thus is a non-symmetrical matrix with $m \times m$ dimensions where a typical element, g_{kl} , indicates the probability of a having a comparative advantage in product k conditional upon having a comparative advantage in product l, based on the information provided in $\boldsymbol{X}$.

The resulting matrix already provides rich information about the probability of developing advantage. However, as argued by Nomaler and Verspagen (2022), we also incorporate information that captures the lack of comparative advantage in a particular product to estimate better the probability that a country has a comparative advantage in another one.

Next, we define the matrix $\boldsymbol{Z} = \boldsymbol{O} - \boldsymbol{X}$, in which $\boldsymbol{O}$ is a matrix consisting entirely of ones and with $m \times n$ dimensions. The elements of the matrix $\boldsymbol{Z}$ thus are defined as follows:

$$z_{ij} = \begin{cases} 1 \; if \; x_{ij} = 0 \\ 0 \; if \; x_{ij} = 1 \end{cases}$$

We define the corresponding conditional probability (product-by-product) matrix $\boldsymbol{H}$ as:

---

[5] The RCA is calculated as:

$$RCA = \frac{E_{ij}/Ej}{Ei/E}$$

where E_{ij} denotes country j exports of product i and the summation over the relevant dimension is indicated by the absence of a subscript. It is also assumed that all countries export at least one product, and all products represent an export of at least one country.

[6] This also corresponds to the vector conceptualized as ubiquity in Hausmann and Hidalgo (2010) where the more countries export a product, the more ubiquitous the product is. Assumedly, higher ubiquity indicates that the capabilities required for producing such a product are more accessible to a large number of countries, and thus, less likely to be of higher complexity.

$$\boldsymbol{H} = (\boldsymbol{X}\boldsymbol{Z}')'/t \; = (\boldsymbol{Z}\boldsymbol{X}')/t$$

where t represents the row-sum of $\boldsymbol{Z}$, i.e., the number of countries that export a given product with no comparative advantage. $\boldsymbol{H}$ is a non-symmetrical matrix with $m \times m$ dimensions where a typical element, denoted as h_{kl}, indicates the probability of having a comparative advantage in product k conditional upon not having comparative advantage in product l, based on the information provided in $\boldsymbol{Z}$. As the following step the two conditional probability matrices are added up and scaled by m (the vector containing the total number of products exported by a given country):

$$\boldsymbol{K} = (\boldsymbol{G} + \boldsymbol{H})/m$$

$\boldsymbol{K}$, therefore, is a matrix of marginal conditional probabilities, with $m \times m$ dimensions. Finally, we obtain a matrix comprised of the estimation of the probabilistic part of the RCA – contained in $\boldsymbol{X}$ - that results from the specialization profile of the country:

$$\boldsymbol{E} = \boldsymbol{X}'\boldsymbol{K}$$

Thus, $\boldsymbol{E}$ is a non-autonomous, (i.e., country-specific) matrix with dimensions $m \times n$ where an element of $\boldsymbol{E}$, denoted as e_{ij}, indicates the probability that country *j* has comparative in product *i* conditional on the information about the whole range of products in which *j* has comparative advantage as well on the information about the range of products in which it does not.

To summarize, the **related variety** probability estimation in $\boldsymbol{E}$, is based on the underlying assumption that if two products, A and B, demand the same capabilities to produce them, these products are related to each other (and likely to be produced by the same country). If B requires capabilities that are very different from capabilities to produce A, these will be unrelated to each other (and unlikely to be produced by the same country), and thus have a lower related variety. Thus, the related variety probability estimation**,** based on the method proposed in Nomaler and Verspagen (2022), accounts for the information which captures similar capabilities, hence the *relatedness*, but also incorporates valuable information captured in the *absence* of those capabilities, which also affect the probability of a country to competitively produce a given product[7] and gain a comparative advantage in the international market.

[7] To illustrate further why this is relevant, Nomaler and Verspagen (2022) show that the absence of specialization frequently coincides with the absence of some other specializations – a kind of *'anti-relatedness'* - which ultimately suggests some sort of competition in specialization.

3.2 Econometric Approach

We start our econometric approach with a modified version of the model proposed in Hausmann and Klinger (2006, 2007), where we employ as explanatory variable the related variety probability estimation described in section 3.1. Following a literature debate, we use 4-year intervals (as opposed to 1-year intervals) to account for the time it takes to develop new products, to diversify production[8]. The resulting equation is then as follows:

$$RCA_{i,c,t+4} = \alpha + \gamma RCA_{\ i,c,t} + \beta E_{i,c,t} + \mu_i + \mu_c + \mu_t + \varepsilon_{i,c,t} \qquad (1)$$

where $RCA_{\ i,c,t}$ is a binary **dependent variable** which captures comparative advantage in product i in country c at the end of a 4-year period; and, $E_{i,c,t}$ is the related variety probability of product i in country c at the beginning of the period[9]. Subsequently, the parameter γ refers to the contribution of having comparative advantage in product i in country c at the beginning of the period to the probability of maintaining such comparative advantage four years later, capturing the persistence of comparative advantage. Likewise, the parameter β captures the effect of related variety on building/keeping comparative advantage at the end of the period. Finally, μ_i, μ_c, and μ_t refer to product, country, and time fixed effects.

Equation (1) estimates the **probability of diversification**: The dependent variable captures whether a given country has a comparative advantage (RCA $\geq$ 1) in a given product of any sort, i.e., extractive and non-extractive products. To compare how diversification differs among different goods (i.e., non-extractive and extractive), we include a second specification where the dependent variable represents if a country has comparative advantage (RCA $\geq$ 1) in a given non-extractive product. For this, the sample is restricted to non-extractive products. A third specification considers a dependent variable that captures if a country has comparative advantage (RCA $\geq$ 1) in extractive commodities. For the latter, the sample is restricted to energy, metals, and minerals commodities[10].

To distinguish how relatedness measures impact upon the probability of gaining advantage in a new product from the impact upon maintaining comparative advantage (or preventing abandonment) in goods already produced*,* we expand equation (1) following Hausmann and Klinger (2007):

[8] Several studies have opted for 5-year periods for this reason (see, for instance, Alonso & Martín, 2019; Boschma & Frenken, 2009). In particular, Alonso and Martín (Alonso & Martín, 2019) replicate the analysis with 4-year intervals and find no significant difference between the 5-year and 4-year periods. Since the panel is built based on a dataset that extends over 24 years, 4-year periods fit the time period while allowing for a reasonable length of time for product development.

[9] The latter term specifically refers to a typical cell, e_{ij} , contained in the E matrix defined in the previous section.

[10] These includes all mining commodities classified under the HS2 codes 26 and 71 and energy commodities under HS4 codes 2709, 2701 and 2711. Energy products do not include any form of processed product.

$$RCA_{i,c,t+4} = \alpha + \gamma RCA_{i,c,t} + \delta(1 - RCA_{i,c,t}) \times E_{i,c,t} + \vartheta(RCA_{i,c,t}) \times E_{i,c,t} + \mu_i + \mu_c + \mu_t + \varepsilon_{i,c,t} \quad (2)$$

parameters δ and ϑ reveal the effect that related variety would have on gaining comparative advantage in a new product and in maintaining it after the end of 4 years, respectively. The term, $E_{i,c,t}$, is not included independently because it is collinear with the two interaction terms. Finally, we expand Equation (2) to include controls at the national level to account for the macroeconomic conditions and other controls, including commodity prices and real exchange rates that might affect diversification efforts:

$$RCA_{i,c,t+4} = \alpha + \gamma RCA_{i,c,t} + \delta(1 - RCA_{i,c,t}) \times E_{i,c,t} + \vartheta(RCA_{i,c,t}) \times E_{i,c,t} + \theta W_{c,t} + \mu_i + \mu_c + \mu_t + \varepsilon_{i,c,t} \quad (3)$$

where W is a matrix of macroeconomic variables which include a) the log of the country-specific mining price index as developed by Deaton (1999); b) the log of real effective exchange rate (REER) index (2010 = 100)[11]; and c) the log of GDP per capita (constant 2010 US dollars), d) investment as a share of GDP. It also includes two dummies capturing extractive commodity dependence: countries categorized as metal-, ore- and mineral-dependent take the value of 1, and 0 otherwise. Similarly, countries categorized as fuel- and gas-dependent take the value of 1, and 0 otherwise. In this way, a country can be energy-dependent, *or* mining-dependent, *or* not dependent on either type of commodity (there is no overlap among energy and mining dependence dummies).

While a linear probability model may serve as a useful point[12], estimating the model using probit (with a specification analogous to Equations (1) through (3)) offers several advantages given the binary nature of the dependent variable, $RCA_{i,c,t+4}$. In particular, we employ the Chamberlain-Mundlak correlated random effects (CRE) probit model in order to ensure the consistency of parameter estimates when including fixed effects, and to provide a more accurate estimation of the magnitude of the marginal effects (Chamberlain, 1982; Wooldridge, 2010). This model enables control for unobserved heterogeneity in a non-linear framework while accounting for potential correlations between individual-specific effects (in this case, product-specific effects) and observed characteristics, e.g., estimated related variety probability measure. The CRE approach introduces the group-level mean of each of the covariates, $\overline{x_i}$, in the probit specification. Adding $\overline{x_i}$ to control for

[11] This refers to the World Bank's definition of REER: the nominal effective exchange rate (a measure of the value of a currency against a weighted average of several foreign currencies) divided by a price deflator or index of costs.
[12] Previous empirical applications (e.g., Alonso & Martín, 2019; Hausmann & Klinger, 2007) have relied on a linear probability models (LPM) as this approach is less computationally intensive and the maximum likelihood with fixed effects is subject to incidental parameters problems when groups are small yielding inconsistent estimates (Greene, 2004). However, our sample allows for a large number of groups and the correlated random effects probit model circumvents the issue of incidental parameters problem (Wooldridge, 2010, 2019).

unobserved heterogeneity (equivalent to $\mu_i + \mu_c + \mu_t$ as done in Equations (1) to (3)) is intuitive as it allows us to estimate the effect of changing $x_{c,i,t}$ while holding country- and/or product-effects fixed (Wooldridge, 2002). The correlated random effects model is then given by:

$$P(RCA_{i,c,t+4} = 1 | x_{i,c,t}) = \Phi[(\psi + \beta x_{i,c,t} + \xi \overline{x_\iota})(1 + \sigma_a^2)^{-1/2}] \qquad (4)$$

where $x_{i,c,t}$ refers to a vector of observable variables at the product- and country-level described in equations (1) to (3), $\overline{x_\iota}$ is the group-level mean (i.e., country and/or product) of each of these variables[13]; and σ_a^2 is the variance for the part of the random effects not captured by the averages $\overline{x_\iota}$. Year, and energy and mining dependence dummies are included in $x_{i,c,t}$ but excluded in $\bar{x}_\iota$. Note that in this setup, if $\xi = 0$ we would obtain the traditional random effects probit model.

This CRE model is our preferred specification and its analogous specification for Equations (1) to (3) are reported in the results section. However, for comparisons, we also include linear probability models based on the basic framework by Hausmann and Klinger (2007) in the Annex[14]. We also run the model specifications separately for the $RCA_{i,c,t+4}$ of all products, $RCA_{i,c,t+4}$ for non-extractive products, and $RCA_{i,c,t+4}$for extractive commodities. In all specifications, standard errors are clustered at the country level.

3.3 Data

To calculate RCAs and related variety measures described in Section 2, we employ bilateral trade data from the BACI 2021 dataset that covers the 1995-2019 period with data collected for more than 5000 products and 220 countries. The BACI 2021 database constructed by CEPII is directly based on UN Comtrade data; it reconciles exporter and importer declarations and defines products at the 6-digit level from the Harmonized System (HS) nomenclature.

---

[13] The CRE specification in equation (4) incorporates a multi-way fixed effect approach which corresponds to the specifications in the LPM model. For this we employ product- and country-level mean terms (where group-level means are generated separately). Time-effects are incorporated in the model by including year dummy variables. In particular, we follow the routine suggested in the Chamberlain RE pooled MLE model described in Wooldridge (2010).

[14] Table 5 in Annex reports the marginal effects of the LPM and CRE probit model in Equation (1) where different fixed effects are used: first, year, country and product effects, and then, product-time and country-time effects (as done in Klinger, (2006) in an LPM setting). In Table 6 and Table 7 the results for all coefficients/marginal effects are presented for Equation (1) and (2) using LPM and CRE probit model also using fixed effects. Results are comparable and remain robust through all specifications. Yet LPM coefficient values tend to be higher.

The price index is calculated using price data from the major extractive commodities[15] extracted from the World Bank's Pink Sheet; commodity trade data from Thibault Fally's dataset[16], and GDP data from the World Development Indicators. The real exchange index (REER), governance effectiveness index, and GDP per capita data were obtained from the World Development Indicators database.

The commodity dependence binary variables were built upon the corresponding categorisation in UNCTAD (2019).

Table 1: Summary Statistics

	N	**Av.**	**SD**	**Min**	**Max**
Related variety	2,958,320	0.02	0.02	-0.06	0.16
RCA	2,958,320	0.19	0.39	0.00	1.00
Non-extractives RCA	2,910,735	0.19	0.39	0.00	1.00
Extractives RCA	47,585	0.23	0.42	0.00	1.00
Country-specific Mining Price Index (log)	2,676,055	0.11	0.25	0.00	1.48
REER Index (log)	1,699,518	4.58	0.14	4.03	5.73
Mining Commodity Dependence	2,676,055	0.10	0.30	0.00	1.00
Energy Commodity Dependence	2,676,055	0.14	0.35	0.00	1.00
Log GDP p.c. (Constant 2010 US$)	2,676,055	9.01	1.41	5.26	11.64
Log of Investment % of GDP	2,568,498	2.71	0.35	0.48	4.69

Table 1 summarises the data employed in the analysis. About 20% of products (in general and for the non-extractive category) were exported with a comparative advantage (i.e., an RCA equal or above to 1). In the case of extractive products, this is slightly higher, as 23% of exports showed comparative advantage.

4. Results

The estimates of equation (1) and its analogous probit specification are presented in Table 2 in Models (1) to (3). Results indicate that having a comparative advantage (RCAi,c,t) at the beginning of a period is a strong

[15] This includes the following commodities and their corresponding HS4 codes: coal (2701), crude oil (2709), gas (2711); Aluminum(2606); Copper (2603); Iron ore (2601); Lead (2601); Nickel (2604); Tin(2609); Zinc (2608); Gold (7108); Silver (7106); and Platinum (7110).

[16] Thibault Fally's database also relies on the BACI database; yet it uses the HS-1992 nomenclature in order to cover a longer period, i.e. from 1995 to 2014 (Fally & Sayre, 2018).

predictor of having a comparative advantage four years later. The estimate on the $RCA_{i,c,t}$ variable is positive and significant at the 1% level. The estimates indicate that having a comparative advantage in a given product at the beginning of a period increases the probability of having it four years later by 28.5 percentage points in the case of all products (Model 1), by 28.4 percentage points in non-extractive products (Model 2); and by 34.0 in extractive products (Model 3). These results remain robust throughout the different specifications presented in Table 2. Similarly, results show that the related variety probability estimate is positive and highly significant. An increase of a standard deviation (0.02) in the related variety estimate increases the probability of (all products') diversification four years later by 6.3 percentage points, (i.e., 3.16*0.02*100) (Model 1); in non-extractive products by 6.3 percentage points (Model 2); and, in extractive products by 7.0 percentage points (Model 3).

Still on Table 2, the estimates of equation (2) are presented in models (4) to (6). The effect of related variety stays positive and highly significant (i.e. at the 1% level); yet the estimated coefficients reveal that its effect on maintaining comparative advantage, $(RCA_{i,c,t}) * E_{i,c,t}$, is higher than on developing new products, $(1 - RCA_{i,c,t}) * E_{i,c,t}$. Specifically, an increase of one standard deviation in the related variety estimate, raises the probability of gaining comparative advantage in a new product (all products category) four years later by 5.8 percentage points, (Model 4); in new non-extractive products by 5.7 (Model 5); and in new extractive commodities by 8.0 percentage points (Model 6)

The results indicate that path dependence may play a bigger role in extractives' diversification than in non-extractives – probably because, on average, the latter requires a more complex and/or diverse set of capabilities.

Furthermore, an increment of 0.02 (a standard deviation) in the related variety estimate increases the predicted probability of maintaining comparative advantage in products (all products category) four years later by 7.2 percentage points (Model 4); and in non-extractive products by 7.2 (Model 5). For extractives, this change would be equivalent to an increment of 6.0 percentage points (Model 6). This suggests that for extractive commodities, path dependence has a stronger effect on 'developing' new (extractive) products vis-à-vis non-extractive products, but it also has a weaker effect on preventing abandonment[17].

[17] To test whether related variety coefficients are statistically different for non-extractive products than for extractive products, we carried out additional regressions in a pooled sample using the LPM approach in which the terms *Related variety,* $E_{i,c,t}$, $(1- RCA_{i,c,t})* E_{i,c,t}$ and $(RCA_{i,c,t})* E_{i,c,t}$, are included, plus their respective interactions with a dummy variable that captures whether if the product is either a mineral, metal, or energy commodity. The results are shown in Table 8 in Annex.

Table 2. Results – Basic Estimation

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES ($RCA_{i,c,t+4}$)	All products	Non-extractive	Extractive	All products	Non-extractive	Extractive
$RCA_{i,c,t}$	0.285***	0.284***	0.340***	0.268***	0.266***	0.353***
	(0.00436)	(0.00439)	(0.00691)	(0.00568)	(0.00577)	(0.00765)
Related variety, $E_{i,c,t}$	3.163***	3.138***	3.512***			
	(0.186)	(0.187)	(0.300)			
$(1- RCA_{i,c,t})^* E_{i,c,t}$				2.915***	2.869***	3.975***
				(0.212)	(0.213)	(0.334)
$(RCA_{i,c,t})^* E_{i,c,t}$				3.612***	3.606***	3.026***
				(0.229)	(0.231)	(0.350)
Observations	2,958,320	2,910,735	47,585	2,958,320	2,910,735	47,585
Country Clusters	228	228	228	228	228	228

*** p<0.01, ** p<0.05, * p<0.1. Note: Country-clustered standard errors are shown in parentheses. All models refer to the CRE probit estimation; coefficients refer to average marginal effects.

In Table 3, Models (4)-(6) represent equation (3) incorporating macroeconomic controls, i.e., the log of the mining price index, the log of the real exchange rate, and log of GDP per capita. The related variety effect on diversification, $E_{i,c,t}$, in models (1), (3), and (5) in Table 3 remains positive and significant at the 1% level. However, the **size of the effect is now smaller** than observed in Table 2. Now, a standard deviation increase (0.02) in related variety is associated with an increase in the probability of diversification of 5.1 percentage points (Model 1), in non-extractive products by 5.0 percentage points (Model 3); and in extractive commodities by 5.2 percentage points (Model 5).

Similarly, the effect of related variety on introducing a new product and maintaining comparative advantage remains positive and highly significant but the effects have reduced regardless of the type of product, as seen in Models (2), (4), and (6). An increase of one standard deviation (0.02) in related variety is associated with a 4.8 percentage point increase in the probability of diversification into non-extractive products after four years (Model 4) and a 7.0 percentage point increase in extractive products (Model 6). Yet related variety has a stronger role in preventing abandonment in non-extractives than in extractives – as earlier observed. The above further underlines that developing comparative advantage in new non-extractive goods is less path-dependent than in mining and energy commodities; in other words, diversifying into non-mining or energy products requires greater efforts for countries specialised in extractive sectors.

We also identify significant differences when examining macroeconomic variables. The coefficient for the mining price index in Models (3) and (4) indicates that a one standard deviation increase (0.25) in the log of the price index is significantly associated with a 12 percentage point decrease in the probability of developing a comparative advantage in non-extractive products four years later (i.e., 0.48*0.25*100). A similar effect is observed across all products (Models 1 and 2), which is also significant at the 1% level. However, no significant effect is found for extractive products.

Moreover, the level of economic development shows a negative association with diversification overall. Models 1 to 4 suggest that an increase of 1.4 (a standard deviation in the sample) in the log of GDP per capita is associated with a reduction in the probability of diversification for all products and non-extractives equivalent to 4.2 percentage points (i.e., 0.03*1.4*100), results significant at the 5% and 10% levels respectively. The negative relationship, however, appears to be much larger and robust with extractive products. Models (5) and (6) in Table 3 indicate that an increment of 1.4 in the log of GDP per capita is associated with a reduction in the probability of having a comparative advantage in extractive commodities equivalent to 13.3 to 13.6 percentage points, significant at the 1% level.

The real exchange rate (REER) does not appear to be significant at any level across these specifications. This is consistent with the previous empirical works that failed to find a relationship between diversification and exchange rates. A possible explanation could be the vast number of currency management regimes and the circular causal relationship which was discussed in the literature review.

Finally, the introduction of controls did not have a noticeable effect on the marginal effects for the initial comparative advantage variable, $RCA_{i,c,t}$ – unlike the related variety marginal effects which became smaller. For this, the introduction of relevant macroeconomic variables linked to the macroeconomic environment is crucial to have a clearer picture of diversification determinants beyond path dependency. Moreover, results in Table 3 show that if the magnitude of the coefficients is compared – based solely on the variation (standard deviation) across countries, macroeconomic factors may play an equal, or stronger, role in explaining different diversification outcomes.

Table 3. Results - Estimation with macroeconomic controls

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES ($RCA_{i,c,t+4}$)	All products	All products	Non-extractive	Non-extractive	Extractive	Extractive
$RCA_{i,c,t}$	0.295***	0.290***	0.294***	0.290***	0.336***	0.356***
	(0.005)	(0.007)	(0.005)	(0.007)	(0.009)	(0.011)
Related variety, $E_{i,c,t}$	2.537***		2.514***		2.600***	
	(0.200)		(0.200)		(0.348)	
(1- $RCA_{i,c,t}$)* $E_{i,c,t}$		2.627***		2.403***		3.475***
		(0.234)		(0.224)		(0.387)
($RCA_{i,c,t}$)* $E_{i,c,t}$		2.726***		2.567***		2.099***
		(0.238)		(0.226)		(0.407)
Price Index (log)	-0.470***	-0.472***	-0.479***	-0.478***	-0.198	-0.213
	(???)	(???)	(???)	(???)	(???)	(???)
GDP per capita (log)	-0.030**	-0.029**	-0.029*	-0.028*	-0.095***	-0.097***
	(0.015)	(0.014)	(0.015)	(0.015)	(0.022)	(0.022)
REER Index (log)	0.018	0.019	0.018	0.018	-0.011	-0.012
	(0.014)	(0.014)	(0.014)	(0.014)	(0.019)	(0.019)
Observations	1,699,518	1,699,518	1,671,028	1,671,028	28,490	28,490
Country Clusters	92	92	92	92	92	92

*** p<0.01, ** p<0.05, * p<0.1. Note: Country-clustered standard errors are shown in parenthesis. All models refer to the CRE probit estimation with product, country, and year effects; coefficients refer to average marginal effects.

Models in Table 4 incorporate further controls: i.e., mining and energy commodity dependence dummies, and the log of investment as a share of GDP. Results in Table 4 indicate that the related variety effect on having comparative advantage – regardless of the type of products – remains significant at the 1% level. The size of the marginal effect, however, decreases slightly. However, it must be said that in the specifications where the variable for investment is introduced the marginal effects increase again slightly. To illustrate this, a one standard deviation (0.02) increase in related variety would be associated with an increment of diversification in a new product four years later equivalent to 3.5 percentage points (Model 7), and if investment is controlled for, 4.4 percentage points (Model 8). Likewise, the equivalent increase in the probability of diversification in extractives would be 4.5 percentage points (Model 11), and if investment is controlled for, 4.9 percentage points (Model 12) (although investment is not significant in the extractive diversification models). In any case, path dependence in new product diversification appears again to be higher for extractives than for non-extractives, as earlier noted.

Furthermore, mining commodity dependence is negatively associated with having a comparative advantage in the category of all products and non-extractives. Specifically, having mining dependency is associated with a reduction in the predicted probability of diversification in all products equivalent to 1.3 percentage points

(Models 1 and 3) and non-extractive products, equivalent to 1.0-1.5 percentage points (Models 5 to 8) significant at the 10% and 5% level (depending on the specification). Controlling for investment, however, seems to attenuate the effect as can be seen throughout Models 1 to 9; whenever this variable is introduced the effect of mining dependency seems to lose significance (or is significant at a lower significance level), with the marginal effect further shrinking. Results in Table 4 also show that mining commodity dependence and diversification in extractive commodities have a positive and highly significant relationship. Namely, mining dependence is associated with an increment in the probability of having comparative advantage in extractives equivalent to 4.9-5.3 percentage points (Models 9 to 12), significant at the 1% level.

Similarly, energy dependency shows the same pattern although the effect appears somewhat less robust than for mining: being dependent on fossil fuels and other energy products is associated with a decrease in diversification in new products (either in the all products or non-extractive products category) of between 1.3 and 1.4 percentage points, significant at the 10% and 5% level. In the specifications where the investment control is introduced, the negative effect loses significance. Likewise, results in Models 9 to 12 suggest that energy dependence is associated with an increment in the probability of diversification between 1.8 and 2.0 percentage points, significant at the 10% significance level. Recent divergence in the diversification trajectories of different oil countries and the overall trend towards higher mining dependence (UNCTAD, 2019) could explain why in recent years the effect of certain dependence could be now stronger for mining.

The effect of mining prices on non-extractive diversification – while smaller – remains negative and significant, even after controlling for commodity dependence and investment. To illustrate this effect, an increase of a standard deviation (0.25) in the log of the price index is associated with a reduced probability of having a comparative advantage in non-extractive products four years later, equivalent to 1.0-1.3 percentage points (Models 5 to 8), effects significant at the 1% level. Similar effects and significance are found for the specification in which all products are considered. Prices remain insignificant in the specifications for extractive products' diversification.

Once controls for commodity dependence and investment are introduced, the negative relationship between GDP per capita and diversification remains negative but appears less strong. Specifically, results indicate that a one standard deviation increment (1.4) in the log of GDP per capita is associated with a reduction in the probability of diversification for all products and non-extractives of between 1.3 and 1.7 percentage points (Models 1 to 8), significant at the 1% level. The effect for extractives, however, is equivalent to 2.2-2.3 percentage points (Models 9 to 12), also significant at the 1% level. The results again highlight that in advanced countries diversification becomes increasingly difficult to attain but also that these countries are less likely to move into extractive commodities – as earlier mentioned.

Table 4. Results - Estimation with macroeconomic controls, commodity dependence dummies, and investment (CRE Probit)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES (RCA $_{i,c,t+4}$)	All products	All products	All products	All products	Non-extractive	Non-extractive	Non-extractive	Non-extractive	Extractive	Extractive	Extractive	Extractive
RCA$_{i,c,t}$	0.295***	0.293***	0.281***	0.284***	0.294***	0.292***	0.279***	0.282***	0.358***	0.353***	0.369***	0.366***
	(0.003)	(0.003)	(0.004)	(0.005)	(0.003)	(0.003)	(0.004)	(0.005)	(0.007)	(0.006)	(0.007)	(0.007)
Related variety, E$_{i,c,t}$	2.097***	2.211***			2.088***	2.203***			1.708***	1.814***		
	(0.113)	(0.114)			(0.113)	(0.114)			(0.296)	(0.296)		
(1- RCA$_{i,c,t}$)* E$_{i,c,t}$			1.784***	2.065***			1.754***	2.039***			2.269***	2.450***
			(0.185)	(0.176)			(0.185)	(0.176)			(0.386)	(0.387)
(RCA$_{i,c,t}$)* E$_{i,c,t}$			2.386***	2.402***			2.394***	2.409***			1.321***	1.378***
			(0.126)	(0.133)			(0.127)	(0.133)			(0.308)	(0.307)
Price Index (log)	-0.050***	-0.042***	-0.051***	-0.041***	-0.052***	-0.043***	-0.053***	-0.043***	0.009	0.010	0.009	0.011
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.014)	(0.015)	(0.014)	(0.015)
GDP per capita (log)	-0.010***	-0.012***	-0.010***	-0.012***	-0.009***	-0.011***	-0.010***	-0.012***	-0.016***	-0.015***	-0.016***	-0.015***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)	(0.004)
Mining dependence	-0.013**	-0.009	-0.013**	-0.009	-0.014**	-0.010*	-0.015**	-0.011*	0.049***	0.053***	0.049***	0.053***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.015)	(0.015)	(0.015)	(0.015)
Energy dependence	-0.013*	-0.010	-0.013**	-0.010	-0.013*	-0.010	-0.014*	-0.011	0.018	0.020*	0.018*	0.020*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.011)	(0.011)	(0.011)	(0.011)
Investment % of GDP (log)		0.013**		0.013**		0.013**		0.013**		-0.005		-0.006
		(0.006)		(0.006)		(0.006)		(0.006)		(0.013)		(0.014)
Observations	2,676,055	2,568,498	2,676,055	2,568,498	2,632,720	2,526,537	2,632,720	2,526,537	43,335	41,961	43,335	41,961
Country Clusters	178	165	178	165	178	165	178	165	178	165	178	165

*** p<0.01, ** p<0.05, * p<0.1. Note: Country-clustered standard errors are shown in parenthesis. All models refer to the CRE probit estimation with product, and year effects; coefficients refer to average marginal effects.

Finally, investment is positively associated with diversification in the all-products and non-extractive products models. Specifically, an increase of one standard deviation (0.35) in the log of the share of investment as GDP is associated with an increment in the probability of diversification equivalent to 0.455 percentage points, significant at the 1% level (Models 6 and 9). Results fail to find the same effect for extractive products, suggesting that, on average, countries with higher levels of investment are less likely to develop towards extractive commodity sectors (perhaps deliberately) – akin to the dynamic observed for more advanced economies.

Estimations based on Table 4 (i.e., Models 3-4, 7-8, and 11-12) were also carried out with additional macroeconomic controls, i.e., log of inflation (from World Development Indicators), and a proxy to account for the quality of institutions, i.e., government effectiveness index (World Governance Indicators). These, however, were not significant in any of the models. Also, to test whether the relationship between product diversification and economic development, i.e., log of GDP per capita, follows a non-linear function, its squared term was introduced in the estimation of models reported in Table 4. The significance of this coefficient was not remarkably high (10%), yet the coefficients indicate a potential nonlinear relationship between GDP per capita and diversification. Namely, this relationship suggests – as highlighted in previous studies (i.e., Hesse, 2008; Klinger & Lederman, 2006) – that, at lower levels of development, export diversification increases but after a certain high-income point it begins to decline. Including these controls did not change much the significance and/or size of the estimated coefficients reported. Results of the above estimations are found in Annex (See Table 9).

4.1 Discussion of Results

A few observations can summarise our results: The related variety measure we use in our analysis (Nomaler & Verspagen, 2022) is a strong predictor of diversification. Our results confirm that path dependence, proxied by this measure, does play a role in predicting what countries produce with comparative advantage and what they do not. Specifically, our results show that this measure is weaker in developing comparative advantage in non-extractive products vis-à-vis extractive products. This suggests that diversifying in non-extractives requires somewhat "bigger jumps" due to more diverse and (probably complex) productive capabilities requirements.

However, related variety does not reveal much about the underlying determinants and macroeconomic incentives facilitating (or hampering) diversification efforts. Results in the previous section show that the effect of related variety is affected by the inclusion of macroeconomic variables (e.g., international prices and investment). It also impacts diversification across sectors differently (in this case, extractive sectors vs other sectors). Likewise, the magnitude of the marginal effects (if the standard deviation in the sample is considered) shows that macroeconomic factors play an important role in explaining differences. Our results support the idea that while path dependence exists, it is far from deterministic. Diversification seems to hinge upon a whole range of macroeconomic factors that ultimately shape the incentives which lead to differences in diversification patterns. In this study, a few are identified and discussed.

Firstly, extractive commodity prices (captured by the country-specific mining index) show a consistent negative association with product diversification in non-extractive products. If extractive commodity dependence and investment are controlled for, the effect of commodity prices on diversification – although smaller – remains negative and significant. This is consistent with previous studies which have highlighted the negative relationship between commodity price shocks and export diversification (i.e., Agosin et al., 2012). Results also show that mining price indices, however, do not incentivize diversification into other non-extractives. Higher prices, thus, may incentivize extracting more of a commodity but are not necessarily conducive to new extractive sectors probably because of the exogenous nature of these resources (i.e., a country either has lithium or not). Additionally, higher prices may not be sufficient to offset the high barriers and requirements involved in developing a new extractive sector.

Likewise, energy- and mining-dependent countries (especially the latter) are less likely to diversify into non-extractive commodity products. Since the effect seems to be particularly strong for mining products, this finding partially contradicts previous studies that indicate that only oil hampers diversification (e.g., Ahmadov, 2014). Possibly this is because while the export concentration in energy-dependent countries remains high, there have been a few mixed experiences more recently.[18]

Yet, in this regard, results suggest that investment can attenuate commodity dependence effects on diversification as investment is positively associated with diversification in non-extractive sectors (and not with extractive commodities). This finding supports the view that diversification is an endogenous process stemming from investments (e.g., Acemoglu and Zilibotti, 1997) as well as previous empirical works (e.g. Esanov, 2012).

Results do not show that the real exchange rate index is statistically associated with diversification (or the lack thereof). The lack of a clear empirical relationship of currency movements with diversification could be attributed not only to the potential bi-directional causality between the variables but also because of the current diversity in exchange rate regimes.

We further confirm – once commodity dependence is controlled for – that at lower levels of development – proxied by GDP per capita – there is more room for diversification, regardless of the type of product considered. However, results also suggest that the more developed a country is, the less likely it will be to diversify into (mining and energy) commodities.

Finally, our results remain robust across estimations in which other controls, such as inflation and governance effectiveness, are included.

5. Conclusion

In this study, we investigated the determinants of export diversification in resource-dependent economies by integrating a novel measure of product relatedness with key macroeconomic variables. We developed an alternative measure of related variety based on the framework proposed by Nomaler and Verspagen (2022), which we argue better captures the complexities of product-level diversification. Using a dataset covering over 5,000 products across multiple countries from 1995 to 2019, we analysed how product relatedness and

[18] Energy-dependent countries like Oman, Trinidad and Tobago, and Qatar became more diversified between 1995- 2017. In contrast, others, such as Azerbaijan, Venezuela, and Nigeria, experienced increased concentration in their economies. (UNCTAD, 2019).

macroeconomic factors—such as commodity prices, exchange rates, and levels of economic development—affect diversification patterns, particularly in extractive versus non-extractive sectors.

Our findings reveal that product relatedness is a strong predictor of diversification, especially within extractive industries, where path dependence is particularly pronounced. This suggests that economies heavily reliant on extractive sectors face significant challenges in diversifying into non-extractive products, requiring greater effort to break away from entrenched specialisation. Furthermore, macroeconomic conditions play a decisive role in shaping diversification outcomes. High commodity prices tend to reinforce specialisation in extractive industries, while lower levels of economic development are associated with more opportunities for diversification into non-extractive sectors. Interestingly, the real exchange rate did not emerge as a significant factor in diversification, though economic development levels showed a clear negative correlation with diversification, particularly in extractive industries.

These results highlight the critical importance of considering product-specific capabilities and broader economic conditions when assessing a country's potential for diversification. For policymakers, this underscores the need for targeted investments and strategic management of macroeconomic variables to promote diversification, particularly in resource-dependent economies. Future research should further explore the interaction between these factors, especially in light of global economic shifts and the increasing focus on sustainable development.

Bibliography

Acemoglu, D., & Zilibotti, F. (1997). Was Prometheus Unbound by Chance? Risk, Diversification, and Growth. *Source: The Journal of Political Economy*, *105*(4), 709–751.

ADB. (2021). *The Greater Mekong Subregion 2030 and Beyond*. https://doi.org/10.22617/TCS210015-2

Agosin, M. R., Alvarez, R., & Bravo-Ortega, C. (2012). Determinants of Export Diversification Around the World: 1962-2000. *World Economy*, *35*(3), 295–315. https://doi.org/10.1111/j.1467-9701.2011.01395.x

Ahmadov, A. K. (2014). *Blocking the Pathway Out of the Resource Curse: What Hinders Diversification in Resource-Rich Developing Countries?* https://www.princeton.edu/~pcglobal/conferences/GLF/ahmadov.pdf.

Alonso, J. A., & Martín, V. (2019). Product relatedness and economic diversification at the regional level in two emerging economies: Mexico and Brazil. *Regional Studies*, *53*(12), 1710–1722. https://doi.org/10.1080/00343404.2019.1605441

Alsharif, N., Bhattacharyya, S., & Intartaglia, M. (2017). Economic diversification in resource rich countries: History, state of knowledge and research agenda. *Resources Policy*, *52*, 154–164. https://doi.org/10.1016/J.RESOURPOL.2017.02.007

Bahar, D., & Santos, M. (2016). *Natural resources and export concentration: on the most likely casualties of Dutch disease* (Center for International Development at Harvard University).

Balassa, B. (1965). Trade Liberalisation and "Revealed" Comparative Advantage1. *The Manchester School*, *33*(2), 99–123. https://doi.org/10.1111/J.1467-9957.1965.TB00050.X

Boschma, R., & Capone, G. (2015). Institutions and diversification: Related versus unrelated diversification in a varieties of capitalism framework. *Research Policy*, *44*(10), 1902–1914. https://doi.org/10.1016/j.respol.2015.06.013

Boschma, R., Minondo, A., & Navarro, M. (2012). *The emergence of new industries at the regional level in Spain A proximity approach based on product-relatedness* (12.01; Papers in Evolutionary Economic Geography). https://doi.org/10.1093/jeg/lbs010

Botta, A., Porcile, G., Spinola, D., & Yajima, G. T. (2023). Financial integration, productive development and fiscal policy space in developing countries. *Structural Change and Economic Dynamics*, *66*, 175-188.

Bresser-Pereira, L. C. (2020). New Developmentalism: development macroeconomics for middle-income countries. *Cambridge Journal of Economics*, *44*(3), 629-646.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, *18*(1), 5–46. https://doi.org/10.1016/0304-4076(82)90094-X

Cimoli, M., Lima, G. T., & Porcile, G. (2016). The production structure, exchange rate preferences and the short-run—Medium-run macrodynamics. *Structural Change and Economic Dynamics*, *37*, 13-26.

Content, J., & Frenken, K. (2016). Related variety and economic development: a literature review. *Http://Dx.Doi.Org/10.1080/09654313.2016.1246517*, *24*(12), 2097–2112.

Corden, W. M., & Neary, J. P. (1982). Booming Sector and De-Industrialisation in a Small Open Economy. *The Economic Journal*, *92*(368), 825. https://doi.org/10.2307/2232670

Deaton, A. (1999). Commodity prices and growth in Africa. *Journal of Economic Perspectives*, *13*(3), 23–40. https://doi.org/10.1257/jep.13.3.23

Diao, X., McMillan, M., & Rodrik, D. (2019). *The recent growth boom in developing economies: A structural-change perspective* (pp. 281-334). Springer International Publishing.

Dosi, G., Pavitt, K., & Soete, L. (1990). The Economics of Technical Change and International Trade. *LEM Book Series*. https://ideas.repec.org/b/ssa/lembks/dosietal-1990.html

Erten, B., & Ocampo, J. A. (2021). The future of commodity prices and the pandemic-driven global recession: Evidence from 150 years of data. *World Development*, *137*, 105164.

Esanov, A. (2012). *Economic Diversification: Dynamics, Determinants and Policy Implications*. https://resourcegovernance.org/sites/default/files/RWI_Economic_Diversification.pdf

Fally, T., & Sayre, J. (2018). *Data on Commodity Trade - Thibault Fally*. https://are.berkeley.edu/~fally/data.html

Guo, Q., & He, C. (2017). Production space and regional industrial evolution in China. *GeoJournal*, *82*(2), 379–396. https://doi.org/10.1007/s10708-015-9689-4

Guzman, M., Ocampo, J. A., & Stiglitz, J. E. (2018). Real exchange rate policies for economic development. *World development*, *110*, 51-62.

Hausmann, R., & Hidalgo, C. A. (2010). *Country Diversification, Product Ubiquity, and Economic Divergence Faculty Research Working Paper Series Ricardo Hausmann* (No. 201; Center for International Development). www.hks.harvard.edu

Hausmann, R., Hwang, J., Rodrik, D., & Kennedy, J. F. (2005). *What You Export Matters*. http://www.nber.org/papers/w11905

Hausmann, R., & Klinger, B. (2006). *Structural Transformation and Patterns of Comparative Advantage in the Product Space* (No. 128; CID Working Papers).

Hausmann, R., & Klinger, B. (2007). *The Structure of the Product Space and the Evolution of Comparative Advantage* (No. 146; CID Working Papers). http://www.cid.harvard.edu/cidwp/128.htm.

He, C., Yan, Y., & Rigby, D. (2018). Regional industrial evolution in China. *Papers in Regional Science*, *97*(2), 173–198. https://doi.org/10.1111/PIRS.12246

He, C., & Zhu, S. (2018). Evolution of Export Product Space in China: Technological Relatedness, National/Local Governance and Regional Industrial Diversification. *Tijdschrift Voor Economische En Sociale Geografie*, *109*(4), 575–593. https://doi.org/10.1111/TESG.12309

Hesse, H. (2008). *Export Diversification and Economic Growth* (No. 21; Commission on Growth and Development).

Hidalgo, C. A., Winger, B., Barabási, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, *317*(5837), 482–487. https://doi.org/10.1126/science.1144581

Klinger, B., & Lederman, D. (2006). *Diversification, Innovation, and Imitation inside the Global Technological Frontier** (No. 3872; Policy Research Working Paper). http://econ.worldbank.org.wps3872

Krugman, P. (1987). The narrow moving band, the Dutch disease, and the competitive consequences of Mrs. Thatcher. Notes on trade in the presence of dynamic scale economies. *Journal of Development Economics*, *27*(1–2), 41–55. https://doi.org/10.1016/0304-3878(87)90005-8

Lall, S. (1992). Technological capabilities and industrialization. *World Development*, *20*(2), 165–186. https://doi.org/10.1016/0305-750X(92)90097-F

Lederman, D., & Maloney, W. F. (2007). Trade Structure and Growth. In D. Lederman & W. F. Manoley (Eds.), *Natural Resources Neither Curse nor Destiny* (pp. 15–40). The World Bank, Stanford University Press.

Maskell, P., & Malmberg, A. (1999). Localised learning and industrial competitiveness. *Cambridge Journal of Economics*, *23*(2), 167–185. https://doi.org/10.1093/CJE/23.2.167

McMillan, M., Rodrik, D., & Verduzco-Gallo, Í. (2014). Globalization, Structural Change, and Productivity Growth, with an Update on Africa. *World Development*, *63*, 11–32. https://doi.org/10.1016/j.worlddev.2013.10.012

Neffke, F., Henning, M., & Boschma, R. (2011). How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions. *Economic Geography*, *87*(3), 237–265. https://doi.org/10.1111/J.1944-8287.2011.01121.X

Nomaler, Ö., & Verspagen, B. (2022). *Some new views on product space and related diversification* (#2022-011; UNU-MERIT Working Paper Series).

Nomaler, Ö., & Verspagen, B. (2024). Related or unrelated diversification: What is smart specialization?. *Structural Change and Economic Dynamics*.

Pérez, C. (2001). Technological change and opportunities for development as a moving target. *CEPAL Review*, 75, 109–130.

Porcile, G., Spinola, D., & Yajima, G. (2023). Growth trajectories and political economy in a Structuralist open economy model. *Review of Keynesian Economics*, *11*(3), 350-376.

Rodrik, D. (2008). *The Real Exchange Rate and Economic Growth* (Brookings Papers on Economic Activity). Sachs and Warner.

Ross, M. L. (2019). What do we know about export diversification in oil-producing countries? *The Extractive Industries and Society*, *6*(3), 792–806. https://doi.org/10.1016/J.EXIS.2019.06.004

Sekkat, K. (2016). Exchange rate misalignment and export diversification in developing countries. *Quarterly Review of Economics and Finance*, *59*, 1–14. https://doi.org/10.1016/j.qref.2015.08.001

Tran, T. A. D., Phi, M. H., & Diaw, D. (2017). Export diversification and real exchange rate in emerging Latin America and Asia: A South–North vs. South-South decomposition. *Journal of International Trade and Economic Development*, *26*(6), 649–676. https://doi.org/10.1080/09638199.2017.1286680

UNCTAD. (2019). *The State of Commodity Dependence 2019*.

van der Ploeg, F., & Poelhekke, S. (2009). Volatility and the natural resource curse. *Oxford Economic Papers*, *61*(4), 727–760. https://doi.org/10.1093/oep/gpp027

Ville, S., & Wicken, O. (2012). The Dynamics of Resource-Based Economic Development : Evidence from Australia and Norway. *Department of Economics, University of Wollongong*, *Working Paper 04-12*, 55. http://ro.uow.edu.au/commwkpapers/241

Wiig, A., & Kolstad, I. (2012). If diversification is good, why don't countries diversify more? The political economy of diversification in resource-rich countries. *Energy Policy*, *40*(1), 196–203. https://doi.org/10.1016/j.enpol.2011.09.041

Winter, S. G., & Nelson, R. R. (1982). *An Evolutionary Theory of Economic Change*. https://papers.ssrn.com/abstract=1496211

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). MIT Press.

Wooldridge, J. M. (2019). Correlated random effects models with unbalanced panels. *Journal of Econometrics*, *211*(1), 137–150. https://doi.org/10.1016/J.JECONOM.2018.12.010

Annex

Annex 1: Additional regressions

Table 5. Comparison of Marginal Effects for Related Variety based on Equation (1)

	(1)	(2)	(3)	(4)
	LPM	CRE Probit	LPM	CRE Probit
	All products RCA $_{i,c,t+4}$	All products RCA $_{i,c,t+4}$	All products RCA $_{i,c,t+4}$	All products RCA $_{i,c,t+4}$
Related variety, $E_{i,c,t}$	4.913***	3.163***	6.859***	4.539***
	(0.301)	(0.186)	(0.391)	(0.0478)
Year	Yes	Yes	-	-
Country	Yes	Yes	-	-
Product	Yes	Yes	-	-
Country*Year	-	-	Yes	Yes
Product*Year	-	-	Yes	Yes
N	2,958,319	2,958,320	2,957,792	2,958,320

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Models 2 and 4 report average marginal effects. Country-clustered SEs are shown in parenthesis.

Table 6. Results – Equation 1: CRE Probit and LPM with Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	LMP	*CRE Probit*	*LMP*	*CRE Probit*	*LMP*	*CRE Probit*
VARIABLES	All products RCA $_{i,c,t+4}$	All products RCA $_{i,c,t+4}$	Non-extractive RCA $_{i,c,t+4}$	Non-extractive RCA $_{i,c,t+4}$	Extractive RCA $_{i,c,t+4}$	Extractive RCA $_{i,c,t+4}$
$RCA_{i,c,t}$	0.540***	0.285***	0.539***	0.284***	0.553***	0.340***
	(0.0108)	(0.00436)	(0.0109)	(0.00439)	(0.0110)	(0.00691)
Related variety, $E_{i,c,t}$	4.913***	3.163***	4.911***	3.138***	3.656***	3.512***
	(0.301)	(0.186)	(0.303)	(0.187)	(0.360)	(0.300)
Observations	2,958,319	2,958,320	2,910,734	2,910,735	47,585	47,585
Adj./Pseudo R-squared	0.411	0.344	0.411	0.345	0.396	0.315
Country Clusters	228	228	228	228	228	228

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Note: Country-clustered SEs are shown in parenthesis. Coefficients are reported for LMP with fixed effects and average marginal effects reported for CRE Probit. All models include year, product and country effects

Table 7. Results – Equation 2: CRE Probit and LPM with Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	LMP FE	CRE Probit	LMP FE	CRE Probit	LMP FE	CRE Probit
VARIABLES	All products RCA $_{i,c,t+4}$	All products RCA $_{i,c,t+4}$	Non-extractive RCA $_{i,c,t+4}$	Non-extractive RCA $_{i,c,t+4}$	Extractive RCA $_{i,c,t+4}$	Extractive RCA $_{i,c,t+4}$
$RCA_{i,c,t}$	0.494***	0.268***	0.491***	0.266***	0.563***	0.353***
	(0.0119)	(0.00568)	(0.0121)	(0.00577)	(0.0126)	(0.00765)
(1- $RCA_{i,c,t}$)* $E_{i,c,t}$	3.643***	2.915***	3.578***	2.869***	4.027***	3.975***
	(0.301)	(0.212)	(0.302)	(0.213)	(0.395)	(0.334)
($RCA_{i,c,t}$)* $E_{i,c,t}$	5.680***	3.612***	5.702***	3.606***	3.286***	3.026***
	(0.362)	(0.229)	(0.367)	(0.231)	(0.461)	(0.350)
Observations	2,958,319	2,958,320	2,910,734	2,910,735	47,585	47,585
Adj./Pseudo R-squared	0.412	0.3461	0.412	0.3468	0.396	0.3170
Country Clusters	228	228	228	228	228	228

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Note: Country-clustered SEs are shown in parenthesis. Coefficients are reported for LMP with fixed effects and average marginal effects reported for CRE Probit. All models include year, product and country effects

Table 8. Statistical difference between commodities and non-commodity products

	(1)	(2)
	LPM	LMP
	All products RCA $_{i,c,t+4}$	All products RCA $_{i,c,t+4}$
$RCA_{i,c,t}$	0.539***	0.491***
	(0.011)	(0.012)
$RCA_{i,c,t}$ * Extractive Commodity Dummy	0.044***	0.107***
	(0.011)	(0.012)
(1- $RCA_{i,c}$)* $E_{i,c}$		3.620***
		(0.300)
($RCA_{i,c}$)* $E_{i,c}$		5.727***
		(0.367)
(1- $RCA_{i,c}$)* $E_{i,c}$ * Extractive Commodity Dummy		-0.564*
		(0.294)
($RCA_{i,c}$)* $E_{i,}$ * Extractive Commodity Dummy		-3.421***
		(0.442)
$E_{i,c,t}$	4.929***	
	(0.302)	
$E_{i,c,t}$ * Extractive Commodity Dummy	-1.838***	
	(0.304)	
Constant	0.022***	0.038***
	(0.003)	(0.004)
N	2,958,319	2,958,319
R-squared	0.412	0.413
Country Clusters	228	228

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Note: Country-clustered SEs are shown in parenthesis. All models include product, country, and year-specific fixed effects.

Table 9 . More controls based on Table 4: Governance effectiveness, inflation and log of GDP per capita² (CRE probit)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All products RCA $_{i,c,t+4}$	All products RCA $_{i,c,t+4}$	All products RCA $_{i,c,t+4}$	Non-extractive RCA $_{i,c,t+4}$	Non-extractive RCA $_{i,c,t+4}$	Non-extractive RCA$_{i,c,t+4}$	Extractive RCA $_{i,c,t+4}$	Extractive RCA $_{i,c,t+4}$	Extractive RCA $_{i,c,t+4}$
RCA$_{i,c,t}$	0.284***	0.285***	0.286***	0.282***	0.283***	0.284***	0.366***	0.368***	0.367***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)
(1-RCA)* E$_{i,c,t}$	2.081***	2.050***	2.093***	2.055***	2.024***	2.065***	2.446***	2.399***	2.522***
	(0.175)	(0.179)	(0.179)	(0.174)	(0.179)	(0.179)	(0.386)	(0.378)	(0.393)
(RCA)* E$_{i,c,t}$	2.413***	2.376***	2.384***	2.420***	2.384***	2.392***	1.376***	1.292***	1.334***
	(0.133)	(0.135)	(0.134)	(0.134)	(0.135)	(0.135)	(0.302)	(0.304)	(0.302)
Price Index (log)	-0.039***	-0.041***	-0.040***	-0.041***	-0.043***	-0.042***	0.013	0.011	0.011
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.016)	(0.016)	(0.015)
GDP per capita (log)	-0.014***	-0.039**	-0.012***	-0.014***	-0.038**	-0.012***	-0.017**	-0.095***	-0.015***
	(0.003)	(0.015)	(0.002)	(0.003)	(0.015)	(0.002)	(0.007)	(0.032)	(0.004)
Mining dependence	-0.009	-0.012*	-0.011*	-0.011*	-0.014**	-0.013**	0.052***	0.044***	0.052***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.015)	(0.015)	(0.016)
Energy dependence	-0.009	-0.010	-0.011	-0.009	-0.011	-0.011	0.022*	0.021*	0.023*
	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.012)	(0.012)	(0.012)
Governance effectiveness	0.003			0.003			0.004		
	(0.004)			(0.004)			(0.008)		
Investment % of GDP(log)	0.013**	0.013**	0.012**	0.013**	0.014**	0.012**	-0.005	-0.004	-0.009
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.014)	(0.013)	(0.015)
GDP per capita² (log)		0.002*			0.001*			0.005**	
		(0.001)			(0.001)			(0.002)	
Inflation (log)			-0.003			-0.003			-0.007
			(0.004)			(0.004)			(0.008)
Observations	2565851	2568498	2478424	2523911	2526537	2437826	41940	41961	40598

*** p<0.01, ** p<0.05, * p<0.10. Note: Country-clustered SEs are shown in parenthesis. Coefficients reported refer to marginal effects. All models include controls for product and year-specific effects.