Optimizing Economic Complexity

Viktor Stojkoski[1,2] and César A. Hidalgo[1,3*]

[1]Center for Collective Learning, CIAS, Corvinus University of Budapest, Budapest, Hungary
[2] Ss. Cyril and Methodius University in Skopje, Skopje, North Macedonia
[3]Center for Collective Learning, IAST, Toulouse School of Economics, Université de Toulouse Capitole, Toulouse, France

Abstract

Efforts to apply economic complexity to identify diversification opportunities often rely on diagrams comparing the relatedness and complexity of products, technologies, or industries. Yet, the use of these diagrams, is not based on empirical or theoretical evidence supporting some notion of optimality. Here, we introduce a method to identify diversification opportunities based on the minimization of a cost function that captures the constraints imposed by an economy's pattern of specialization and show that this *ECI optimization* algorithm produces recommendations that are substantially different from those obtained using relatedness-complexity diagrams. This method advances the use of economic complexity methods to explore questions of strategic diversification.

Keywords: economic complexity**,** economic development, policy
JEL Codes: O11, O25, C61

Introduction

To achieve sustainable economic growth economies must adapt to changes in markets and technologies.[1,2] Changes in economic structure, however, involve strategic considerations such as identifying diversification opportunities while facing the constraints imposed by an economy's existing productive structure.[3–5]

During the last couple of decades, economic complexity has become a common method to explore questions of strategic diversification.[6–30] The use of these methods is supported by the notion that economic complexity predicts economic growth[6,9,13,14,31], and thus, connects changes in an

---

[*] Corresponding author: cesar.hidalgo@tse-fr.eu

economy's productive structure to its growth potential. Yet, the practical implementation of these ideas often relies on the use of *relatedness-complexity diagram*s: a graphical introduced more than a decade ago[9] to visually inspect an economy's diversification opportunities.[9,32,33]

In a relatedness-complexity diagram, economic activities, such as product exports, industries, or technologies, are presented in a two-way scatter plot (Figure 1) with relatedness in the x-axis and complexity on the y-axis. These diagrams are usually divided into quadrants to help facilitate their interpretation.
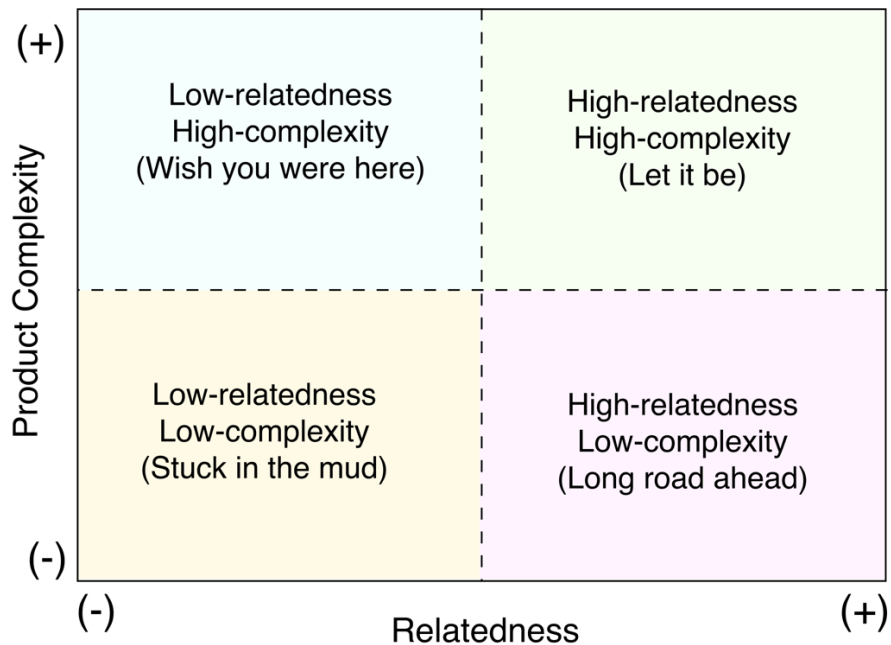

Figure 1. Illustrative example of a relatedness-complexity diagram.

The relatedness axis provides a measure of affinity between a location and an activity.[7,10–12,33–36] This is an estimate of how compatible the activity is with an economy's current productive structure (as expressed by the presence of other activities) or how "easy" it is for that economy to enter that activity (e.g. specialize its exports or employment in pharmaceuticals).[*]

---

[*] Technically this is a simplification, since a proper estimate of entry or exit probabilities requires defining a full model including more than relatedness type variables. For instance, when estimating changes in bilateral trade patterns at the product level, Jun et al. used a model with a total of fifteen different factors, including three types of relatedness.[35]

Complexity provides a measure of the value of each potential activity.[6,37] This is because estimates of a product's complexity are correlated with higher levels of income, and thus, serve as estimates of an activity's contribution to an economy's growth potential.

These diagrams are used to identify activities in the productive frontier, which are those that are relatively high in complexity and relatedness. These are activities that are valuable and feasible for an economy, making a relatedness-complexity diagram a simple and intuitive tool to reason about strategic diversification (an idea similar to an efficient frontier in formal economic models of portfolio optimization[38–40]).

Yet, because these are diagrams based on proxy measures, and do not consider higher order effects (such as future changes in the relatedness and complexity of activities), they do not guarantee that the activities selected through them are optimal according to a more formal criterion.

Here, we address this gap by introducing an optimization method to identify diversification opportunities while considering the path dependencies implied by an economy's existing productive structure. We show that this optimization procedure results in recommendations that are substantially different from those obtained from visually inspecting relatedness-complexity diagrams and that are better at leveraging an economy's existing capacities. This represents an improvement over the use of relatedness-complexity diagrams for economic diversification strategy.

The Use of Economic Complexity in Strategic Diversification

Economic complexity techniques have become common in strategic diversification efforts because of their ability to explain structural transformation processes and quantify the potential value of an economy's pattern of specialization.[6,31,37,41] Today, economic complexity ideas are part of the official industrial strategy of diversified middle-income economies, like Malaysia[42] , rich natural resource dependent economies, such as Saudi Arabia[43], and small economies, like Armenia[44]. These methods have also been applied to guide foreign direct investment efforts in Mexico[45] and have been used to justify the creation of national economic data observatories in Brazil, Mexico, and Spain among other places.[46,47] In Europe, economic complexity tools are commonly used in

reports focused on regional innovation and smart diversification strategies[33,48,49] and were used to compare the relative position of Europe, the United States, and China in key technologies in an influential report prepared by a team led by Mario Draghi.[50]

Most of these efforts bring economic complexity ideas into practice by using *relatedness-complexity* or *diversification frontier* diagrams (Figure 1). These diagrams combine a measure of affinity, such as relatedness, with a measure of value, such as complexity, to identify activities that are both, compatible with an economy's current patterns of specialization (high relatedness) and attractive based on their ability to generate sustainable growth (high complexity).

But the use of these diagrams carries important limitations.

First, relatedness is an incomplete measure of economic potential. In fact, an economy's probability of entering or exiting an activity is better approximated by models that include relatedness as one of many variables. For instance, a model of economic potential that includes also terms for an economy's specialization in each activity, trends in that specialization, and/or multiple forms of relatedness (e.g. relative relatedness[51], bilateral relatedness[35], or industry- and occupation-specific types of relatedness[10,52]). These models are significantly better in terms of predictive accuracy than a model using relatedness alone. Relatedness still provides valuable information in these models, since it captures the importance of spillovers among activities that are hard to capture with other variables (for instance, changes in an economy's exports of cars that can be attributed to that economy's specialization in engines, trucks, wheels, etc.). That's why the academic literature has developed several models that combine measures of relatedness with dozens of other predictors to estimate an economy's potential to enter a specific activity[10,35,53–60]. The applied literature, however, still often uses relatedness in isolation[32,33,49]—which can be illustrative for educational purposes—but mistakes a key component of an estimate of potential (relatedness) with estimates of potential (the full model).

Second, while complexity and relatedness change over time, it is common for applied work to only use current values. For instance, by drawing diagrams using static values for the complexity of an activity instead of taking into consideration how that complexity might change over time. In some

cases, changes can be substantial. Between 1962 and 2016 the product complexity ranking of fully assembled cars (SITC 7810)—according to the product complexity index (PCI)—fell from 17 to 208 (out of 817 activities). This is not because fully assembled cars became less sophisticated, but because many middle- and lower-income economies, such as Thailand, Morocco, and Turkey, became specialized exporters of them. In this car assembly example, a middle- or lower-income country looking to enter the car industry in the late 1980s would overestimate the increase in complexity associated with entering that sector and would underestimate the competition of other entrants.

Third, theoretical work on economic complexity[61,62] has shown that optimal diversification strategies require balancing a portfolio of related and unrelated diversification opportunities. The balance of this portfolio is expected to change with an economy's level of complexity in the form of an inverted U-shape, meaning that unrelated diversification attempts are more beneficial at an intermediate level of relatedness. Yet, while recent work[32,51] has acknowledged the need to create a balanced portfolio of related and unrelated activities, there is still no principled way to estimate that portfolio. Here we show that the ECI optimization method naturally leads to a combination of related and unrelated diversification targets that matches the behavior expected for an optimal portfolio.

And fourth, it is important not to confuse the use of relatedness-complexity diagrams, or of an ECI optimization algorithm, with a method to select sectors. This is a more universal limitation that is important to note at the outset. Economic complexity methods are a useful tool for industrial policy, not because they represent an "ideal" way to select sectors, but because they are a powerful way to dispel sectors chosen through less principled approaches. In the world of international development, it is common for economies that are wildly different to receive sectoral advice motivated by global trends regardless of an economy's local capacities (e.g. AI, Big Data, Sustainable Energy, Biotech). While this type of advice can be soothing for political elites, and even sound when there are available "windows of opportunity,"[63] it can also lead to pie in the sky development efforts that fail to take into account local conditions. By providing a principled way to estimate an economy's probability of success in each sector, together with a measure of its potential value, economic complexity methods provide a rough map that can be used to ground

overly optimistic targets (more details about the use of economic complexity methods in practice can be found in[32]).

Here, we address the first three limitations by introducing an optimization method that helps move the practical use of economic complexity methods beyond the use of heuristics and the visual inspection of relatedness-complexity diagrams. Our approach involves defining a target level of economic complexity or economic growth and using an optimization method to identify a portfolio of new activities that minimizes a measure of the "effort" required to enter them. The estimate of effort is calculated as the required increase in comparative advantage needed to enter an activity that must be realized at an intermediate point in time or "steppingstone." The optimization method also includes a forward-looking model that we use to derive a future Product Complexity Index (PCI) for products, incorporating the second line of criticism.

Armed with this model, we then explore the properties of this optimization procedure using international trade data by country and employment by industry data for cities in the United States. We find that ECI optimization suggests activities that are better at leveraging an economy's comparative advantage, while adjusting the level of relatedness according to its stage of development. It also identifies a more balanced portfolio of related and unrelated activities that the ones identified using relatedness complexity diagrams. In sum, the *ECI optimization* method advances the practical use of economic complexity by providing a quantitatively rigorous approach for the support of economic development strategies.

Results

The ECI Optimization Method

The *ECI optimization* process begins by defining a target level of economic complexity which we operationalize by either selecting directly a value for the economic complexity index (ECI) or by choosing that value indirectly by inverting the empirically observed relationship between growth and ECI. That is, for an economy with an economic complexity of $ECI_c(t)$ at time t, we define a

target $ECI_c^*(t + \Delta t)$. Since ECI is defined[6] as the average complexity (PCI) of the activities an economy specializes in,[6] we are looking for a solution to:

$$ECI_c^* = \frac{1}{M_c^*} \sum_p M_{cp}^* PCI_p$$

Where ECI_c^* is our target level of economic complexity, M_{cp}^* is the future specialization matrix that we are looking to optimize (our unknown), $M_c^* = \sum_p M_{cp}^*$ is the number of activities in which economy c specializes in after the optimization has taken place, and PCI_p are the predicted product complexity indexes of the activities at time $(t + \Delta t)$. For the standard definitions of ECI and PCI see the following review paper[37].

The key mathematical question here is how to determine the specialization matrix M_{cp}^*. That requires defining an optimization criterion or cost function. Here we look for the M_{cp}^* that achieves the target level of economic complexity (ECI^*) while minimizing the sum of the increase in comparative advantages at an intermediate time point, or "steppingstone." In principle, we could choose other constraints, such as minimizing the added volume of exports in the case of trade or minimizing the increase in added payroll for industries. For illustration and clarity purposes, we minimize the sum of comparative advantages as our criterion, since it provides reasonable results and avoids some limitations of other optimization constraints. For instance, minimizing the added volume of employment or exports biases the optimization process towards smaller activities.

As usual, we define M_{cp}^* as a binarized measure of a specialization matrix R_{cp}^*, which is an estimate of revealed comparative advantage[*] (RCA),

$$R_{cp} = \frac{X_{cp}X}{X_c X_p}$$

---

[*] $R_{cp} = X_{cp}X \Big/ X_c X_p$, where X_{cp} is an output matrix, summarizing the output of location c in activity p, and $X_c = \sum_p X_{cp}$, $X_p = \sum_c X_{cp}$, and $X = \sum_{c,p} X_{cp}$.

where X_{cp} is a measure of output, volume, or value added (depending on availability, e.g. total exports of country c in product p, total payroll paid by city c in industry p, etc.). Also, we use Einstein's notation where muted indexes were added over (e.g. $X_c = \sum_p X_{cp}$).

That is, $M_{cp}^* = 1$ if $R_{cp}^* \geq 1$ and 0 otherwise. We then calibrate the coefficients of our forecast model by using a linear regression of the form:

$$r_{cp}(t + \Delta t) = b_1 r_{cp}(t + \tau) + b_2 r_{cp}(t) + b_3 \omega_{cp}(t) + b_4 \widetilde{\omega}_{cp}(t) + b_0 + e_{cp}(t) \qquad (2)$$

where

$$r_{cp}(t) = \log(R_{cp}(t) + 1)$$

and $R_{cp}(t)$ is the specialization (RCA), t is the initial time point, $t + \Delta t$ is the final time point, and $\tau < \Delta t$ is the steppingstone time point. Also, $\omega_{cp}(t)$ is the relatedness of location c in activity p, $\widetilde{\omega}_{cp}(t)$ is the relative relatedness of the same location-activity pair[51], and $e_{cp}(t)$ is the error term.

We calibrate the coefficients in equation (2) using historical data, by assuming that coefficients are different for entry models (defined as $R_{cp}(t) < 1 \,\&\, R_{cp}(t + \Delta t) \geq 1$) and exit models (defined as $R_{cp}(t) > 1 \,\&\, R_{cp}(t + \Delta t) \leq 1$). For instance, for a model starting on the year t =2012, with a five-year steppingstone (τ=5) and a ten-year horizon (Δt=10), we calibrate the coefficients by regressing the output in 2022 with a steppingstone in 2017. We justify this calibration by showing that the coefficients determined through this process are similar regardless of the starting year. In fact, as we should expect, the coefficients depend more on the duration of the steppingstone than on the final year. For instance, a 10-year forecast with a 5-year steppingstone has similar coefficients when choosing 2002 or 2012 as the starting year but has different coefficients than a 10-year forecast with a steppingstone on year 2.

In Table 1, we show two examples of regression analyses predicting output in 2022, using a 5-year steppingstone in the upper panel and a 10-year steppingstone in the lower panel. In both cases, we find that the models including all four explanatory variables (column 3 for entry models and

column 6 for exit models) have the highest explanatory power. Additionally, we observe that all explanatory variables show a positive and statistically significant relationship with the output variable.

Table 1. Entry and Exit Model for Target Year 2022

| | *Dependent variable: log(1 + RCA) (2022) [in 5 years]* | | | | | |
| | Entry | | | Exit | | |
	(1)	(2)	(3)	(4)	(5)	(6)
log of RCA (2017)	0.648***		0.636***	0.889***		0.874***
	(0.003)		(0.003)	(0.005)		(0.005)
log of RCA (2012)	0.283***		0.215***	0.007		0.022***
	(0.004)		(0.005)	(0.006)		(0.006)
Relatedness (2012)		0.878***	0.180***		-0.653***	0.284***
		(0.007)	(0.007)		(0.042)	(0.022)
Relative Relatedness (2012)		0.043***	0.014***		0.271***	0.023***
		(0.001)	(0.001)		(0.006)	(0.003)
Observations	134948	134948	134948	24217	24217	24217
R^2	0.459	0.140	0.465	0.754	0.087	0.757
Adjusted R^2	0.459	0.140	0.465	0.754	0.087	0.757

| | *Dependent variable: log(1 + RCA) (2022) [in 10 years]* | | | | | |
| | Entry | | | Exit | | |
	(1)	(2)	(3)	(4)	(5)	(6)
log of RCA (2012)	0.641***		0.627***	0.823***		0.817***
	(0.003)		(0.003)	(0.006)		(0.006)
log of RCA (2002)	0.284***		0.196***	0.025***		0.020***
	(0.005)		(0.006)	(0.006)		(0.007)
Relatedness (2002)		1.056***	0.249***		-0.506***	0.044
		(0.010)	(0.009)		(0.048)	(0.031)
Relative Relatedness (2002)		0.038***	0.014***		0.227***	0.031***
		(0.001)	(0.001)		(0.006)	(0.004)
Observations	107669	107669	107669	21121	21121	21121
R^2	0.417	0.126	0.423	0.660	0.062	0.661
Adjusted R^2	0.417	0.126	0.423	0.660	0.062	0.661

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Since there are multiple choices for a starting year, given a steppingstone and a horizon, we average the model's estimates across all possible initial years t, yielding a more reliable parameter set (see the Supplementary Material). Using these coefficients, we set up a second equation where we

allow the comparative advantages in the steppingstone year to change from those in the initial year by an amount equal to an unknown matrix W_{cp}. That is, we set up the model:

$$r_{cp}(t + \Delta t) = b_1 \log(R_{cp}(t) + W_{cp} + 1) + b_2\, \mathrm{r}_{cp}(t) + b_3 \omega_{cp}(\mathrm{t}) + b_4 \widetilde{\omega}_{cp}(\mathrm{t}) + b_0 + e_{cp}(t) \quad (3)$$

which we can solve algebraically for W_{cp} when setting $R_{cp}(t + \Delta t) = 1$. That is, we estimate the W_{cp} that correspond to an entry, making W_{cp} an estimate of the change in specialization (the added *RCA*) that an economy needs to achieve by the steppingstone year $(\mathrm{t}+ \tau)$ for the model to predict that this economy will specialize in an activity by the target year $(t + \Delta t)$.

Finally, we use the coefficients in W_{cp} in a binary optimization program (a.k.a. 0-1 integer optimization)[64,65] . The program minimizes the sum of the weights in W_{cp} for new specializations $(M_{cp}(t) = 0 \rightarrow M_{cp}(t + \Delta t) = 1)$ that do not require a boost from W_{cp}, subject to reaching an economic complexity equal or larger than the target. That is, we look for:

$$\frac{\sum_p M_{cp}(t + \Delta t) PCI_p}{\sum_p M_{cp}(t + \Delta t)} \geq ECI_c^*$$

While minimizing:

$$\min(\sum_p W_{cp} M_{cp}^*(t + \Delta t))$$

Where PCI_p is obtained from the specialization matrix predicted by equation (3) when setting $W_{cp} = 0$ for time $t + \tau$. We justify this assumption by noting that while the *PCI* values would change after the optimization, here we optimize only one economy, which should translate into a minor impact on PCI_p. Thus, we approximate the complexity of activities by using the future *PCI* implied by the model's forecast.

M_{cp}^* provides the portfolio of activities that optimally reaches a target level of economic complexity while minimizing the increase in comparative advantage needed to achieve that target at the intermediate time point.

Notice that our approach assumes that one economy optimizes at a time (no general equilibrium effects with multiple economies optimizing simultaneously while knowing that other economies are also attempting to optimize). This simplifies the problem but does not fully capture general equilibrium effects. Despite this limitation, ECI optimization should enhance the application of economic complexity methods by providing a more strategic and forward-looking approach to diversification.

Figure 2 summarizes the ECI optimization process and its benefits over standard complexity-relatedness diagrams. First, instead of relying solely on relatedness, it incorporates a more detailed and generalizable measure of the effort needed to add an activity, represented by the additional *RCA* required in the steppingstone year in the context of a multivariate model. Second, ECI optimization adopts a forward-looking perspective by integrating forecasts of changes in both RCA and complexity values over time. This dynamic view helps economies avoid overestimating potential complexity gains from industries that may become more ubiquitous as other countries also develop specialization in them.
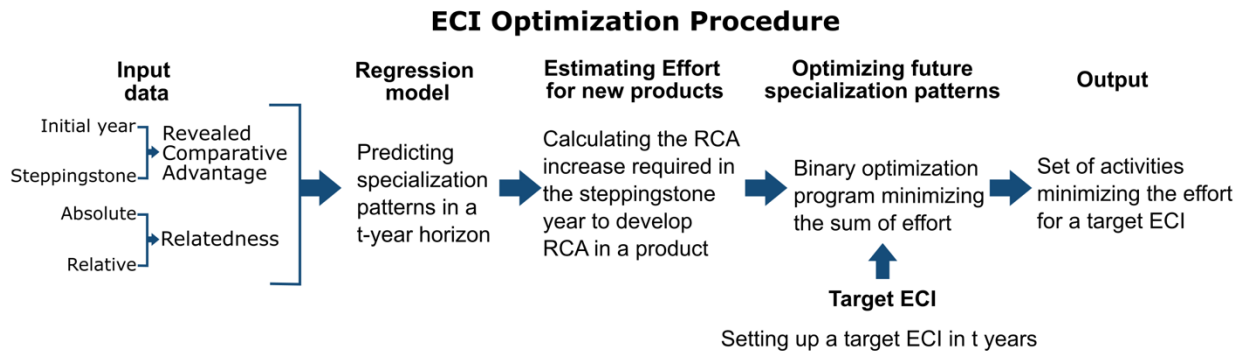

Figure 2. ECI Optimization Systems Diagram.

Example

We illustrate the ECI Optimization method by drawing an effort (W)-complexity (PCI) diagram (Figure 3) for the economy of Vietnam. In this diagram we plot the expected Product Complexity Index (PCI_p) (at the target year) for each potential new activity p, alongside the effort W_{cp} required at the steppingstone year for an economy to reach a "positive" level of specialization ($R_{cp}(t + \Delta t) \geq 1$).

The optimization process begins by selecting activities in the upper-left quadrant (Figure 3 a), which represent the most efficient choices in terms of the tradeoff between complexity and effort. But how many activities do we need to choose to reach an ECI target? And how do we trade-off between low effort and high complexity?

As we raise the target ECI (Figures 3b and 3c), we must gradually include more activities—some with higher required effort (moving rightward) and some with slightly lower complexity (moving downward). This sequential inclusion reflects the trade-off between increasing complexity and managing the additional effort required.
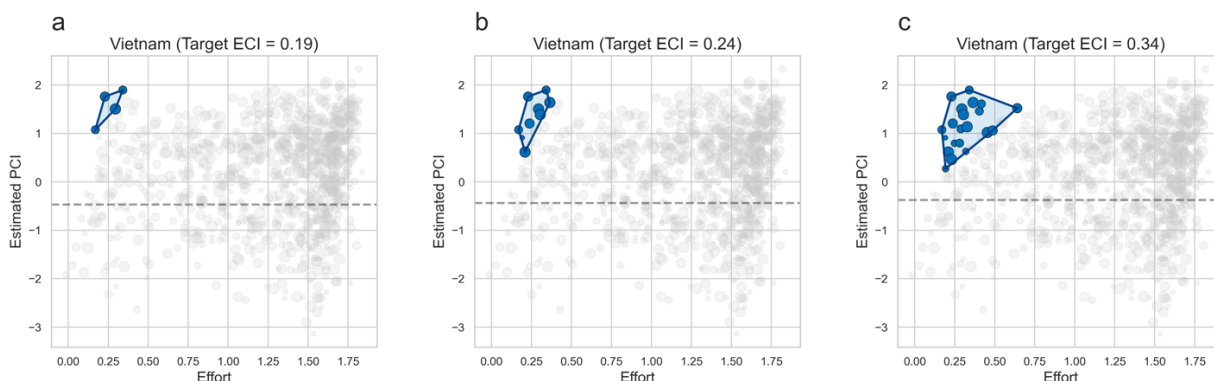

Figure 3. The ECI Optimization Method. a) ECI Optimization effort-complexity diagram for Vietnam's exports to reach a target ECI of 0.19 (starting from an ECI of 0.14). The selected activities to reach a target ECI are highlighted in blue. **b)** Same as **a)**, only for a target ECI of 0.24. **c)** Same as **a)**, only for a target ECI of 0.29.

Properties of ECI Optimization

We explore the properties of the ECI optimization method by comparing it with a benchmark that approximates a common practice in the use of relatedness-complexity diagrams. That is identifying

activities that are relatively high in complexity and relatedness. We implement this benchmark by first min-max normalizing the relatedness and complexity (PCI) of each activity (to a 0 to 1 scale) and then taking the product of these values. Then, we select the N activities with the highest score needed to reach the target ECI.

For this example, we use international trade data disaggregated by the HS4 classification (Revision 1992), downloaded from the Observatory of Economic Complexity (oec.world) and focus on using 2022 data to model the economic structure in 2032. That means setting 2027 as a steppingstone year. For each location, we target an increase in ECI of 0.1 standard deviations ($ECI(t + \Delta t) = ECI(t) + 0.1$)). More details about the data are available in the appendix.

Figure 4 compares ECI optimization with the *relatedness-complexity* diagram benchmark.

Figure 4a looks at the average Revealed Comparative Advantage (RCA) in 2022 for the products suggested by both methods. This shows that *ECI optimization* suggests products where countries currently have higher specialization compared to those suggested by the benchmark. This is a reasonable outcome, since ECI optimization minimizes the additional RCA required to achieve a target ECI. But this is also an important difference between the two methods, as it shows that ECI optimization focuses on activities that leverage existing capacities (as expressed in products close to full levels of specialization (e.g. RCA~0.8)).

Figure 4b compares the two methods by looking at their average relative relatedness (relatedness centered on a country's mean). Compared to the benchmark, ECI optimization selects products that are relatively more unrelated for lower and middle complexity economies. The relationship defined by the ECI optimization choices also has a curvature resembling the one observed for the optimal diversification model introduced by Alshamsi et al.,[61,62] which shows that effective diversification strategies require targeting both related and unrelated activities. This is confirmed by the inset of Figure 4b, which shows that the standard deviation of the recommended activities drastically increases for countries with a medium-high value of ECI (ECI~0.5). Hence, ECI optimization may endogenously lead to a more balanced portfolio of related/unrelated activities in

accordance with models of optimal portfolio diversification, balancing the pursuit of complex products with the feasibility associated with their relatedness.

Figure 4c plots the number of suggested products as a function of an economy's diversity (the number of activities in which a country had RCA>1 in 2022). Here, both methods behave similarly, showing that, while they recommend different activities, they recommend a similar number of them. Still, ECI optimization suggests a slightly lower number of new activities compared to the benchmark model.

Finally, Figure 4d plots the minimum value of additional export volume required to reach comparative advantage in the suggested products as a function of diversity in 2022. This additional export volume can be considered an alternative measure of the effort or cost needed to develop new specializations. We find that the ECI optimization method suggests slightly lower additional volume compared to the benchmark model (around USD 1B on average), suggesting that it effectively balances the trade-off between the feasibility of expanding export volumes and the potential gains in economic complexity.
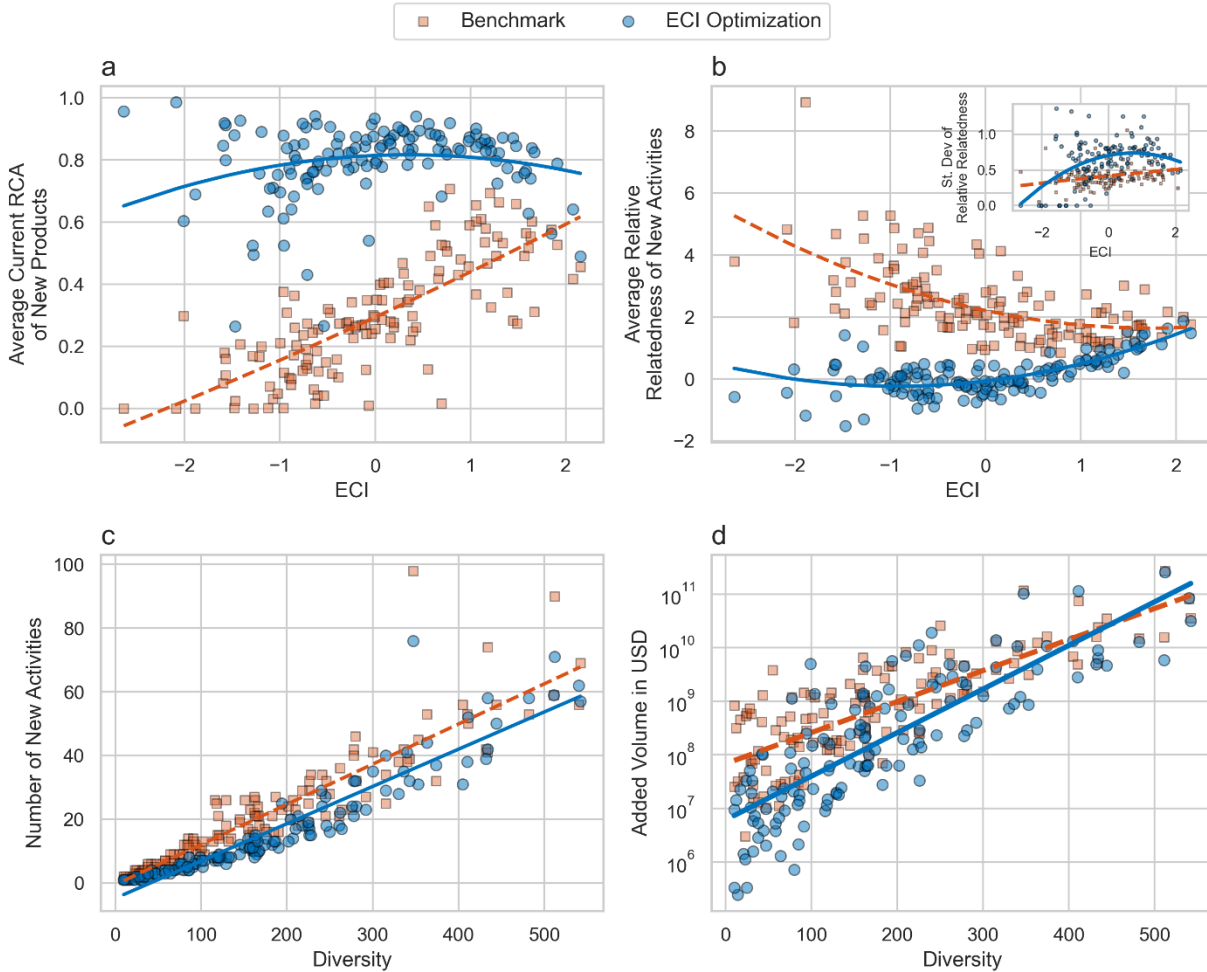
Figure 4. Properties of ECI Optimization in International Trade Data. a) Average RCA (in 2022) of the products suggested by ECI optimization and the benchmark model as a function of a country's ECI in 2022. **b)** Average relative relatedness of the suggested products as a function of a country's ECI in 2022. The inset plot shows the variance of the relative relatedness of the suggested products. **c)** The number of suggested activities as a function of the initial diversity (number of activities in which the country had comparative advantage in 2022). **d)** Estimated added export volume to gain comparative advantage in the suggested products as a function of the initial diversity. **a-d)** For each country we assume an increase of 0.1 of the ECI value in 2022. All scatter charts include a quadratic fit line.

In the supplementary material, we repeat this analysis using the USA's MSA payroll by industry data, finding similar results on the subnational level. Hence, our results are robust to the granularity of the data (national vs. subnational) and the classification of activities (product exports vs. industry payroll). Overall, these results suggest that ECI optimization recommends activities where economies already have some comparative advantage while balancing relatedness with the required level of added output.

Targeting Economic Growth

Finally, we use ECI optimization to target a level of economic growth. We do this by inverting an econometric growth model to ask for the ECI that is compatible with a growth target. We start by constructing the following panel growth regression using ECI as an explanatory variable:

$$\log \frac{GDPpc_{t+10}}{GDPpc_t} = a_1 ECI_{ct} + a_2 \log GDPpc_t + a_3 ECI_{ct} \times \log GDPpc_t + \gamma_t + u_{ct}. \qquad (3)$$

Here, the dependent variable is the annualized 10-year growth rate of GDP per capita (in PPP constant 2021 USD) of a country (we consider two periods 1999–2009 and 2009–2019).[*] Besides ECI, we use two additional explanatory variables. First, we use a z-score normalized value for the log of initial GDP per capita (normalized across our sample of countries for each year), capturing Solow's idea of economic convergence.[66][†] Second, we use the interaction of ECI with the z-score of the initial log of GDP per capita, capturing the idea that the contribution of economic complexity to future economic growth is stronger for lower income economies.[9] We also use period fixed effects in order to account for any omitted variables that vary over the two decades and may impact economic growth.

Once we estimate the parameters of the model (see Table S1 in the Supplementary Information) we can set a target growth rate and invert the equation to find the ECI which is compatible with that level of growth. Here, we use two countries with a similar ECI in 2022 as examples: Thailand (ECI = 0.98) and Mexico (ECI = 0.99).

We start by using our steppingstone model to estimate the export structure of these countries and their corresponding ECI in 2032 without any optimization. Our model estimates that in 2032 Thailand will have an ECI = 0.933, whereas Mexico will have ECI = 0.926, meaning that the

---

[*] We opt to model growth in terms of annualized changes of GDP per capita PPP (constant 2021 USD) because this approach captures cross-country variations in income levels and cost structures, providing a more comparable measure of growth. These predictions can be translated into standard GDP growth rates by incorporating population growth, or into simpler GDP per capita growth rates by discounting the adjustments for PPP.
[†] This z-score transformation helps us account for the non-stationary nature of GDP per capita and provide a consistent prediction.

countries may experience a slight decrease in their *ECI*, remaining largely at a similar level of complexity.

Next, we use our growth regression model to provide an estimate for the expected economic growth rate that supports the predicted ECI. Even though the countries have a similar ECI, they do have a slightly different expected growth rate (assuming that the other economic conditions, will remain the same as in the last time period, e.g. fixed effects and relative GDP per capita): Thailand has an expected annualized growth rate of 3.23%, whereas Mexico has 3.15%. This is because they have a different starting level of GDP per capita.

Then, we ask the question of which is the optimal portfolio according to ECI optimization for these countries to increase their growth potential to 3.5%? To answer this, we invert the growth equation (3) and recover an expected value of ECI that supports the targeted growth. For Thailand, we find that ECI = 1.225, whereas for Mexico ECI = 1.287.

We use these values as target ECI's in ECI optimization and suggest new activities for the countries. Figures 5 a and 5 b depict, respectively, the results for Thailand and Mexico using effort-complexity diagrams. In the figures we also highlight the results from the benchmark model. We observe that ECI optimization yields significantly different predictions compared to those of the benchmark model. In each case, ECI optimization suggests the products located in the top left of the diagram, which are "high gain" and "low effort." By contrast, the benchmark model suggests many more products are located on the top right, meaning that they require high effort.

In Figure 5c, we summarize the ECI Optimization output for Thailand in a table, listing the products in the order of the first target growth rate they recommended, up to the target of 3.5%. For lower targets, the ECI Optimization method suggests products like 8532: Electrical Capacitors and 7320: Iron Springs, which Thailand can develop with relatively low effort. As the target ECI rises to a value that corresponds to an annualized growth rate of 3.5%, the list expands to include higher-effort, higher-complexity products such as 8477: Rubberworking Machinery.

By comparison, the suggested products for Mexico (shown in Figure 5d) differ significantly from those of Thailand. For lower ECI targets, the ECI Optimization method recommends goods like 8547: Metal Insulating Fittings and 8485: Additive Manufacturing Machines, while higher ECI targets introduce additional products such as 8480: Metal Molds into the portfolio. This contrast highlights how the method adapts recommendations to fit the economic contexts of different economies, accounting for each economy's predicted future structure and existing comparative advantages.
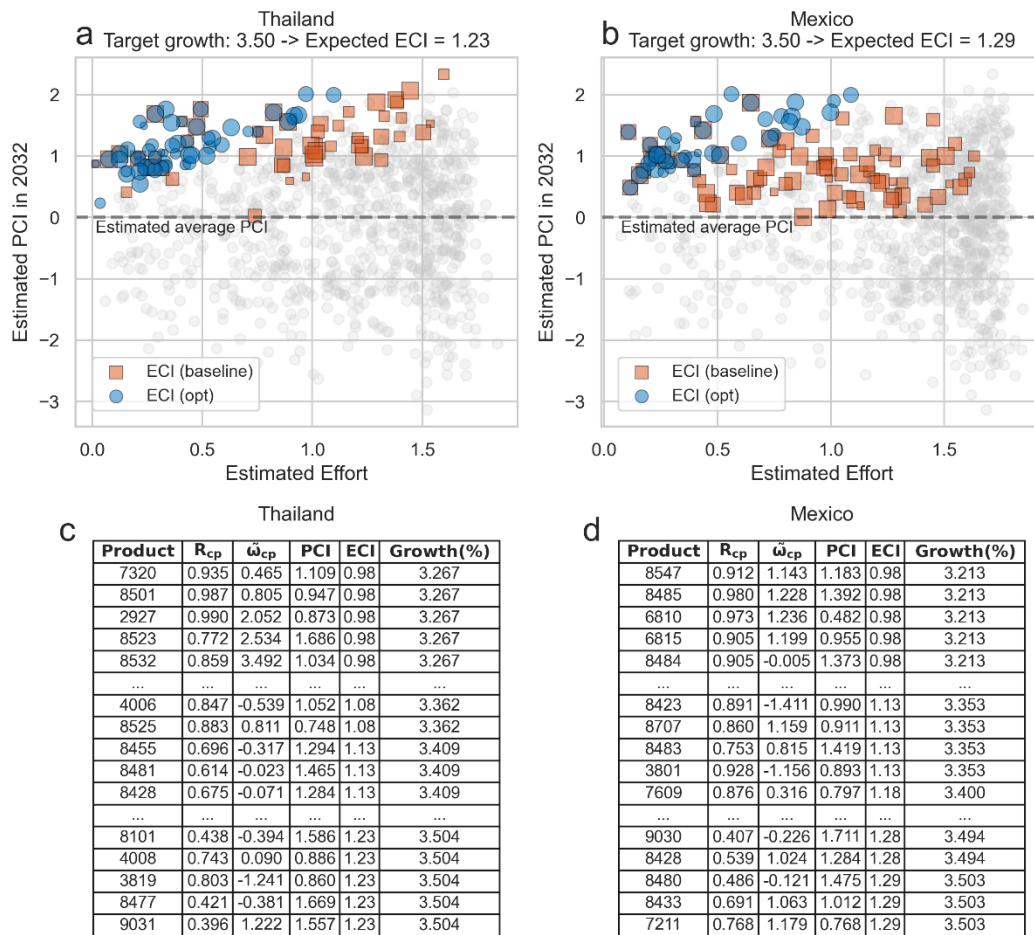

a Thailand — Target growth: 3.50 -> Expected ECI = 1.23

b Mexico — Target growth: 3.50 -> Expected ECI = 1.29

c Thailand

Product	R_{cp}	$\tilde{\omega}_{cp}$	PCI	ECI	Growth(%)
7320	0.935	0.465	1.109	0.98	3.267
8501	0.987	0.805	0.947	0.98	3.267
2927	0.990	2.052	0.873	0.98	3.267
8523	0.772	2.534	1.686	0.98	3.267
8532	0.859	3.492	1.034	0.98	3.267
...	...	...	...	...	...
4006	0.847	-0.539	1.052	1.08	3.362
8525	0.883	0.811	0.748	1.08	3.362
8455	0.696	-0.317	1.294	1.13	3.409
8481	0.614	-0.023	1.465	1.13	3.409
8428	0.675	-0.071	1.284	1.13	3.409
...	...	...	...	...	...
8101	0.438	-0.394	1.586	1.23	3.504
4008	0.743	0.090	0.886	1.23	3.504
3819	0.803	-1.241	0.860	1.23	3.504
8477	0.421	-0.381	1.669	1.23	3.504
9031	0.396	1.222	1.557	1.23	3.504

d Mexico

Product	R_{cp}	$\tilde{\omega}_{cp}$	PCI	ECI	Growth(%)
8547	0.912	1.143	1.183	0.98	3.213
8485	0.980	1.228	1.392	0.98	3.213
6810	0.973	1.236	0.482	0.98	3.213
6815	0.905	1.199	0.955	0.98	3.213
8484	0.905	-0.005	1.373	0.98	3.213
...	...	...	...	...	...
8423	0.891	-1.411	0.990	1.13	3.353
8707	0.860	1.159	0.911	1.13	3.353
8483	0.753	0.815	1.419	1.13	3.353
3801	0.928	-1.156	0.893	1.13	3.353
7609	0.876	0.316	0.797	1.18	3.400
...	...	...	...	...	...
9030	0.407	-0.226	1.711	1.28	3.494
8428	0.539	1.024	1.284	1.28	3.494
8480	0.486	-0.121	1.475	1.29	3.503
8433	0.691	1.063	1.012	1.29	3.503
7211	0.768	1.179	0.768	1.29	3.503

Figure 5. Applying ECI Optimization for Diversification. a) Effort-Complexity diagram for Thailand to increase its 2032 ECI to a value that corresponds of a 3.5% annualized GDP per capita growth rate. **b)** Same as **a**, for Mexico. **c)** List of suggested products for Thailand for a sequential increase of ECI to a value that corresponds to a 3.5% annualized GDP per capita growth rate. The products are ordered by the first target ECI at which they are suggested. **d)** Same as **c**, just for Mexico.

Discussion

Economic complexity applications often rely on straightforward interpretations of relatedness-complexity diagrams, overlooking nuanced optimization aspects—such as the fact that relatedness is only one component of a diversification potential model or that comparative advantages change over time. Without considering these additional dimensions, the recommendations derived from such methods can lead to misleading results.

To address this gap, we developed the *ECI Optimization* method. This method provides a structured approach to strategic diversification by modeling future specialization patterns and using these forecasts to inform an optimization framework. By applying model predictions as weights within a linear program, the ECI Optimization method identifies a portfolio of activities that achieve a target level of economic complexity with minimal cost.

We visualized the ECI Optimization in a cost-complexity diagram, where potential activities are mapped based on their complexity and the cost required to develop comparative advantage in each. The optimization algorithm begins by selecting activities in the top-left of this diagram—those with high complexity and low cost—and gradually incorporates additional activities toward the top-right as higher complexity targets are pursued.

We explored the properties of our method using data from international trade (and the US cities (MSAs) and industries in the SM). We observed that ECI Optimization recommends a portfolio of target activities that naturally balances among related and unrelated activities following what is predicted by the optimal diversification theory of Alshamsi et al.[61]

Finally, we demonstrated the practical application of ECI Optimization by incorporating it in a simple framework that exploits the knowledge of a target economic growth rate in order to suggest activities that optimize the desired outcome. Using Thailand and Mexico as examples (two countries with similar level of ECI), we showed how the method ensures that the diversification strategy provides tailored recommendations for each economy's productive structure.

Altogether, this approach offers a mathematically rigorous alternative to traditional diversification strategies, equipping policymakers with a tool that balances complexity gains with practical feasibility in economic development efforts.

Yet, our approach is not without limitations.

First, although the ECI Optimization method offers a structured approach to enhancing economic complexity, it may render the ECI susceptible to Goodhart's Law: when a measure becomes a target for optimization, it can lose its efficacy as an accurate indicator.[67–69] By explicitly optimizing for ECI, we risk altering the patterns that the index was originally designed to capture. This could lead to a scenario where the ECI no longer reliably reflects an economy's productive structure. As a result, the ECI might lose its effectiveness as a diagnostic tool for assessing economic structure and guiding policy, thereby limiting its utility in promoting sustainable economic development.

A second limitation concerns the general application of economic complexity as a tool for policy. While economic complexity methods are commonly used to suggest activities that could drive economic growth, it is crucial to remember that the index itself is a 'philosophically positive' measure.[32,70] That is, ECI reflects multiple underlying outcomes—knowledge, institutional frameworks, and global connections, among others. Thus, any application of ECI optimization should not stop at the generated recommendations but should also delve into the broader dynamics that shape these outcomes. For example, Economic complexity, as used in Malaysia's New Industrial Master Plan,[42] is a mission acting as a rallying flag hoping to coordinate multiple activities towards a common goal. An effective application of ECI-based strategies requires acknowledging the spatial and organizational aspects of knowledge growth and diffusion that underpin economic structures.[53,71–75] This includes, for instance, fostering connections with global leaders in specific sectors or cultivating institutions that can support these new activities. Without these foundational elements, the benefits of ECI recommendations may be short-lived.

Third, our illustration of the method is based on a limited diversification model using only a few parameters (two measures of density and a measure of specialization) and a simple linear model specification for the stepping stone funciton.[10,35,52,76–80] While this allowed us for easy analytical

formulation of the optimization procedure (in terms of a linear zero-one integer problem), it could be the case that more obscure but more precise machine learning models perform better. Indeed, future efforts to build on this ECI optimization idea could explore the benefits of using more comprehensive diversification models and specifications, including not easily invertible functions (e.g. functions considering multiple steppingstones (separately for specialization and relatedness)).

Fourth, our optimization model primarily considers a single dimension of economic activity—such as international trade data or industrial activities within MSAs. Economic complexity, however, is inherently multidimensional.[31] It encompasses not only the sophistication of products or industries but also the complexity of other structures like technologies,[81,82] research,[48] and the digital economy.[83,84] These additional dimensions are crucial for an economy's potential for inclusive and sustainable growth. By limiting our analysis to a singular dimension, we risk overlooking these critical aspects, thereby restricting the comprehensiveness and applicability of our diversification recommendations.

Finally, a key limitation of this approach is that it does not fully account for general equilibrium effects, where multiple economies optimize simultaneously while considering how others are also optimizing. While RCA and ECI help mitigate this issue by being relative measures, implying that the global average adjusts as countries change their specializations, our model does not explicitly account for interdependencies in decision-making across countries. Addressing this would require a game-theoretic or agent-based extension of the framework, which we leave for future research.

Yet, despite these limitations, *ECI Optimization* advances the policy toolkit of economic complexity by providing a mathematically grounded approach for strategic diversification. As such, it should motivate new research on comprehensive methodologies for strategic diversification.

Methods

Data: We use two datasets to evaluate the performance of ECI Optimization. In the main manuscript we present results using international trade data, disaggregated by the HS4

Classification (Rev. 1992), spanning from 1998 to 2022, and sourced from the Observatory of Economic Complexity (oec.world). To reduce noise from random fluctuations in annual trade values, we preprocess the data using a 4-year moving average to represent exports of each product in a given year (e.g., exports in 2022 are averaged over 2019-2022). Moreover, for each year, we include only countries with exports exceeding USD 1 billion and that had a population of above 1 million. We also exclude products with trade values below 500,000 USD. These criteria help minimize noise from smaller economies or low-trade-volume products.

In the SM, we display results from models using subnational U.S. payroll data by metropolitan area from 2003 to 2022 taken from the United States Census Bureau's County Business Patterns (https://www.census.gov/programs-surveys/cbp.html). Identically to the trade dataset, to reduce variability, we calculate four-year averages for payrolls across years. Also, we further refine this dataset by including only metropolitan areas with total payrolls exceeding 100 thousand USD and activities where the total payrolls surpass 150 thousand USD.

Regression model parameter estimation: We use an Ordinary Least Squares (OLS) approach to estimate the parameters defined in equation (2), applying it separately for each initial year within a fixed timeframe Δt and stepping stone τ. To improve stability and reduce noise from year-to-year fluctuations, we average the OLS estimates across all possible initial years t, yielding a more reliable parameter set.

We thereby emphasize that our estimation differentiates between models of entry and exit. The entry models use data only on activities where $R_{cp}(t) < 1$, i.e., modeling only the activities in which locations could potentially establish new specializations at time $t + \Delta t$. In contrast, the exit models consider only data on activities where $R_{cp}(t) \geq 1$, capturing only those activities from which locations are likely to exit in terms of specialization. This approach is standard in the literature and underscores that the strength of the specialization variables and the relatedness variables differs depending on whether we aim to explain the emergence of new activities or the disappearance of existing ones.

In the empirical analysis presented throughout the manuscript we set the timeframe Δt to 10 years, and the stepping stone τ to 5 years. Also, as a starting point for the future predictions we use the latest year, 2022, meaning that we use our models to predict the geography of economic activities in 2023, and use ECI Optimization to select optimal portfolios for the 2027 steppingstone.

Conflict of Interest Disclosure

César A. Hidalgo is a co-founder of Datawheel LLC a data visualization and distribution company developing oec.world and other economic data visualization projects.

Data Statement

The data and code that can be used to reproduce the results of this manuscript can be found at: https://doi.org/10.7910/DVN/RXDCI1

References

1. Soubbotina, T. P. *Beyond Economic Growth: An Introduction to Sustainable Development*. (World Bank Publications, 2004).

2. Stiglitz, J. E. Towards a New Paradigm for Development: Strategies, Policies and Processes. (2002) doi:10.7916/D8MC98VB.

3. Chang, H. & Andreoni, A. Industrial Policy in the 21st Century. *Development and Change* **51**, 324–351 (2020).

4. Juhász, R. & Lane, N. The political economy of industrial policy. *Journal of Economic Perspectives* **38**, 27–54 (2024).

5. Juhász, R., Lane, N. & Rodrik, D. The New Economics of Industrial Policy. *Annual Review of Economics* **16**, 213–242 (2024).

6. Hidalgo, C. A. & Hausmann, R. The building blocks of economic complexity. *Proceedings of the national academy of sciences* **106**, 10570–10575 (2009).

7. Hidalgo, C. A., Klinger, B., Barabási, A.-L. & Hausmann, R. The product space conditions the development of nations. *Science* **317**, 482–487 (2007).

8. Bustos, S., Gomez, C., Hausmann, R. & Hidalgo, C. A. The dynamics of nestedness predicts the evolution of industrial ecosystems. *PloS one* **7**, e49393 (2012).

9. Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M. & Simoes, A. *The Atlas of Economic Complexity: Mapping Paths to Prosperity*. (Mit Press, 2014).

10. Jara-Figueroa, C., Jun, B., Glaeser, E. L. & Hidalgo, C. A. The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12646–12653 (2018).

11. Neffke, F., Henning, M. & Boschma, R. How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic geography* **87**, 237–265 (2011).

12. Neffke, F. & Henning, M. Skill relatedness and firm diversification. *Strategic Management Journal* **34**, 297–316 (2013).

13. Chávez, J. C., Mosqueda, M. T. & Gómez-Zaldívar, M. Economic complexity and regional growth performance: Evidence from the Mexican Economy. *Review of Regional Studies* **47**, 201–219 (2017).

14. Domini, G. Patterns of specialization and economic complexity through the lens of universal exhibitions, 1855-1900. *Explorations in Economic History* **83**, 101421 (2022).

15. Koch, P. Economic Complexity and Growth: Can value-added exports better explain the link? *Economics Letters* **198**, 109682 (2021).

16. Ourens, G. Can the Method of Reflections help predict future growth? *Documento de Trabajo/FCS-DE; 17/12* (2012).

17. Stojkoski, V. & Kocarev, L. The relationship between growth and economic complexity: evidence from Southeastern and Central Europe. (2017).

18. Stojkoski, V., Utkovski, Z. & Kocarev, L. The impact of services on economic complexity: Service sophistication as route for economic growth. *PloS one* **11**, e0161633 (2016).

19. Barza, R., Jara-Figueroa, C., Hidalgo, C. A. & Viarengo, M. Knowledge intensity and gender wage gaps: evidence from linked employer-employee data. (2020).

20. Basile, R., Cicerone, G. & Iapadre, L. Economic complexity and regional labor productivity distribution: evidence from Italy. *The Review of regional studies* 201–219 (2019).

21. Hartmann, D., Guevara, M. R., Jara-Figueroa, C., Aristarán, M. & Hidalgo, C. A. Linking economic complexity, institutions, and income inequality. *World development* **93**, 75–93 (2017).

22. Can, M. & Gozgor, G. The impact of economic complexity on carbon emissions: evidence from France. *Environ Sci Pollut Res* **24**, 16364–16370 (2017).

23. Dong, F. *et al.* How does digital economy affect carbon emissions? Evidence from global 60 countries. *Sci Total Environ* doi:doi: 10.1016/j.scitotenv.2022.158401.

24. Lapatinas, A. The effect of the Internet on economic sophistication: An empirical analysis. *Economics Letters* **174**, 35–38 (2019).

25. Boleti, E., Garas, A., Kyriakou, A. & Lapatinas, A. Economic complexity and environmental performance: evidence from a world sample. *Environmental Modeling & Assessment* **26**, 251–270 (2021).

26. Neagu. The Link between Economic Complexity and Carbon Emissions in the European Union Countries: A Model Based on the Environmental Kuznets Curve (EKC) Approach. *Sustainability* **11**, 4753 (2019).

27. Romero, J. P. & Gramkow, C. Economic complexity and greenhouse gas emissions. *World Development* **139**, 105317 (2021).

28. Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A. & Pietronero, L. A new metrics for countries' fitness and products' complexity. *Scientific reports* **2**, 1–7 (2012).

29. Cristelli, M., Gabrielli, A., Tacchella, A., Caldarelli, G. & Pietronero, L. Measuring the intangibles: A metrics for the economic complexity of countries and products. *PloS one* **8**, e70726 (2013).

30. Sciarra, C., Chiarotti, G., Ridolfi, L. & Laio, F. Reconciling contrasting views on economic complexity. *Nature communications* **11**, 3352 (2020).

31. Stojkoski, V., Koch, P. & Hidalgo, C. A. Multidimensional economic complexity and inclusive green growth. *Communications Earth & Environment* **4**, 130 (2023).

32. Hidalgo, C. A. The policy implications of economic complexity. *Research Policy* **52**, 104863 (2023).

33. Balland, P.-A., Boschma, R., Crespo, J. & Rigby, D. L. Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional studies* **53**, 1252–1268 (2019).

34. Hidalgo *et al.* The Principle of Relatedness. in *Unifying Themes in Complex Systems IX* (eds. Morales, A. J., Gershenson, C., Braha, D., Minai, A. A. & Bar-Yam, Y.) 451–457 (Springer International Publishing, 2018).

35. Jun, B., Alshamsi, A., Gao, J. & Hidalgo, C. A. Bilateral relatedness: knowledge diffusion and the evolution of bilateral trade. *Journal of Evolutionary Economics* **30**, 247–277 (2020).

36. Boschma, R., Minondo, A. & Navarro, M. The Emergence of New Industries at the Regional Level in Spain: A Proximity Approach Based on Product Relatedness. *Economic Geography* **89**, 29–51 (2013).

37. Hidalgo, C. A. Economic complexity theory and applications. *Nature Reviews Physics* **3**, 92–113 (2021).

38. Markowitz, H. Modern portfolio theory. *Journal of Finance* **7**, 77–91 (1952).

39. Farrell, M. J. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (General)* **120**, 253 (1957).

40. Canton, E. A portfolio approach to research funding. *Research Policy* **54**, 105129 (2025).

41. Balland, P.-A. *et al.* The new paradigm of economic complexity. *Research Policy* **51**, 104450 (2022).

42. Ministry of Investment, Trade and Industry. New Industrial Master Plan for Malaysia. https://www.nimp2030.gov.my/.

43. Saudi Arabia. *United States Department of State* https://www.state.gov/reports/2024-investment-climate-statements/saudi-arabia/.

44. Dept, I. M. F. M. E. and C. A. Republic of Armenia: 2023 Article IV Consultation and Second Review Under the Stand-By Arrangement and Request for Modifications of Performance Criteria and Monetary Policy Consultation Clause-Press Release; Staff Report; and Statement by the Executive Director. *IMF Staff Country Reports* **2023**, (2023).

45. DataMéxico | Data México. https://www.economia.gob.mx/datamexico/en.

46. Observatório DataMPE Brasil. *Observatório DataMPE Brasil* https://datampe.sebrae.com.br.

47. DataViva. https://www.dataviva.info/pt/.

48. Balland, P.-A. & Boschma, R. *Do Scientific Capabilities in Specific Domains Matter for Technological Diversification in European Regions?* (2021).

49. Deegan, J., Broekel, T. & Fitjar, R. D. Searching through the Haystack: The relatedness and complexity of priorities in smart specialization strategies. *Economic Geography* **97**, 497–520 (2021).

50. EU competitiveness: Looking ahead - European Commission. https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en.

51. Pinheiro, F. L., Hartmann, D., Boschma, R. & Hidalgo, C. A. The time and frequency of unrelated diversification. *Research Policy* 104323 (2021).

52. Farinha, T., Balland, P.-A., Morrison, A. & Boschma, R. What drives the geography of jobs in the us? unpacking relatedness. *Industry and Innovation* **26**, 988–1022 (2019).

53. Koch, P., Stojkoski, V. & Hidalgo, C. A. The role of immigrants, emigrants and locals in the historical formation of European knowledge agglomerations. *Regional Studies* **58**, 1659–1673 (2024).

54. Romero, J. P. *et al.* Complexity-based diversification strategies: a new method for ranking promising activities for regional diversification. *Spatial Economic Analysis* 1–24 (2024) doi:10.1080/17421772.2024.2384744.

55. Tacchella, A., Zaccaria, A., Miccheli, M. & Pietronero, L. Relatedness in the era of machine learning. *Chaos, Solitons & Fractals* **176**, 114071 (2023).

56. Fessina, M., Albora, G., Tacchella, A. & Zaccaria, A. Identifying key products to trigger new exports: an explainable machine learning approach. *Journal of Physics: Complexity* **5**, 025003 (2024).

57. Tiits, M., Kalvet, T., Ounoughi, C. & Yahia, S. B. Relatedness and product complexity meet gravity models of international trade. *Journal of Open Innovation: Technology, Market, and Complexity* **10**, 100288 (2024).

58. Rosenow, S. K. & Mealy, P. Turning Risks into Reward: Diversifying the Global Value Chains of Decarbonization Technologies. (2024).

59. Hausmann, R., Stock, D. P. & Yıldırım, M. A. Implied comparative advantage. *Research Policy* 104143 (2021).

60. Hausmann, R., Stock, D. P. & Yıldırım, M. A. Implied comparative advantage. *Research Policy* **51**, 104143 (2022).

61. Alshamsi, A., Pinheiro, F. L. & Hidalgo, C. A. Optimal diversification strategies in the networks of related products and of related research areas. *Nature communications* **9**, 1328 (2018).

62. Waniek, M., Elbassioni, K., Pinheiro, F. L., Hidalgo, C. A. & Alshamsi, A. Computational aspects of optimal strategic network diffusion. *Theoretical Computer Science* **814**, 153–168 (2020).

63. Lee, K. & Malerba, F. Catch-up cycles and changes in industrial leadership:Windows of opportunity and responses of firms and countries in the evolution of sectoral systems. *Research Policy* **46**, 338–351 (2017).

64. Williams, H. P. Integer Programming. in *Logic and Integer Programming* vol. 130 25–70 (Springer US, Boston, MA, 2009).

65. Karp, R. M. Reducibility Among Combinatorial Problems. in *50 Years of Integer Programming 1958-2008* (eds. Jünger, M. et al.) 219–241 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010). doi:10.1007/978-3-540-68279-0_8.

66. Solow, R. M. A Contribution to the Theory of Economic Growth. *Q J Econ* **70**, 65–94 (1956).

67. Goodhart, C. A. E. Problems of Monetary Management: The UK Experience. in *Monetary Theory and Practice* 91–121 (Macmillan Education UK, London, 1984). doi:10.1007/978-1-349-17295-5_4.

68. Chrystal, K. A., Mizen, P. D. & Mizen, P. D. Goodhart's law: its origins, meaning and implications for monetary policy. *Central banking, monetary theory and practice: Essays in honour of Charles Goodhart* **1**, 221–243 (2003).

69. Goodhart, C. *Goodhart's Law*. vol. 227 (Edward Elgar Publishing Cheltenham, UK, 2015).

70. Hidalgo, C. A. Knowledge is non-fungible. in *Routledge International Handbook of Complexity Economics* 50–63 (Routledge, 2024).

71. Jaffe, A. B., Trajtenberg, M. & Henderson, R. Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics* **108**, 577–598 (1993).

72. Cipolla, C. M. The diffusion of innovations in early modern Europe. *Comparative studies in Society and History* **14**, 46–52 (1972).

73. Kerr, S. P., Kerr, W., Özden, Ç. & Parsons, C. High-Skilled Migration and Agglomeration. *Annu. Rev. Econ.* **9**, 201–234 (2017).

74. Trippl, M. & Maier, G. Knowledge Spillover Agents and Regional Development. in *Innovation, Growth and Competitiveness* (eds. Nijkamp, P. & Siedschlag, I.) 91–111 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011). doi:10.1007/978-3-642-14965-8_5.

75. Bahar, D. & Rapoport, H. Migration, knowledge diffusion and the comparative advantage of nations. *The Economic Journal* **128**, F273–F305 (2018).

76. Diodato, D., Neffke, F. & O'Clery, N. Why do industries coagglomerate? How Marshallian externalities differ by industry and have evolved over time. *Journal of Urban Economics* **106**, 1–26 (2018).

77. Bahar, D., Rosenow, S., Stein, E. & Wagner, R. Export take-offs and acceleration: Unpacking cross-sector linkages in the evolution of comparative advantage. *World Development* **117**, 48–60 (2019).

78. Boschma, R. & Capone, G. Institutions and diversification: Related versus unrelated diversification in a varieties of capitalism framework. *Research Policy* **44**, 1902–1914 (2015).

79. Cortinovis, N., Xiao, J., Boschma, R. & van Oort, F. G. Quality of government and social capital as drivers of regional diversification in Europe. *Journal of Economic Geography* **17**, 1179–1208 (2017).

80. Zhu, S., He, C. & Zhou, Y. How to jump further and catch up? Path-breaking in an uneven industry space. *Journal of Economic Geography* **17**, 521–545 (2017).

81. Lybbert, T. J. & Xu, M. Innovation-adjusted economic complexity and growth: Do patent flows reveal enhanced economic capabilities? *Review of Development Economics* **26**, 442–483 (2022).

82. Ivanova, I., Strand, Ø., Kushnir, D. & Leydesdorff, L. Economic and technological complexity: A model study of indicators of knowledge-based innovation systems. *Technological forecasting and social change* **120**, 77–89 (2017).

83. Stojkoski, V., Koch, P., Coll, E. & Hidalgo, C. A. Estimating digital product trade through corporate revenue data. *Nature Communications* **15**, 5262 (2024).

84. Juhász, S., Wachs, J., Kaminski, J. & Hidalgo, C. A. The Software Complexity of Nations. Preprint at http://arxiv.org/abs/2407.13880 (2024).

Supplementary Material for: Optimizing Economic Complexity

Viktor Stojkoski[1,2] and César A. Hidalgo[1,3]

[1]Center for Collective Learning, Corvinus University of Budapest, Budapest, Hungary
[2] Ss. Cyril and Methodius University in Skopje, Skopje, North Macedonia
[3]Toulouse School of Economics, Université de Toulouse Capitole, Toulouse, France

Abstract

Efforts to apply economic complexity to identify diversification opportunities often rely on diagrams comparing the relatedness and complexity of products, technologies, or industries. Yet, the use of these diagrams, is not based on empirical or theoretical evidence supporting some notion of optimality. Here, we introduce a method to identify diversification opportunities based on the minimization of a cost function that captures the constraints imposed by an economy's pattern of specialization and show that this *ECI optimization* algorithm produces recommendations that are substantially different from those obtained using relatedness-complexity diagrams. This method advances the use of economic complexity methods to explore questions of strategic diversification.

Keywords: economic complexity**,** economic development, policy
JEL Codes: O11, O25, C61

Table of Contents

Regression Model Performance in International Trade Data

In Figure S1, we present a heatmap of the average coefficient estimates, their standard errors, and p-values for entry models describing equation (2) in the main manuscript, estimated by using international trade data and varying the timeframe Δt and stepping stone τ. Figure S2 complements this by showing the average coefficient of determination for these models. We find that all coefficients exhibit a positive relationship with future specialization, as expected, and that, on average, they are statistically significant at the 0.01 level. Additionally, the models demonstrate an average coefficient of determination of 0.48, indicating a good fit. The coefficient of determination follows a gradient pattern, with higher values observed when τ is closer to Δt.

This pattern is primarily due to the autoregressive relationship between the RCA stepping stone $R_{cp}(t + \tau)$ and the dependent variable. As τ approaches Δt, the magnitude of the coefficient for $R_{cp}(t + \tau)$ increases, while the other coefficients correspondingly decrease.

Figures S3 and S4 illustrate the same heatmaps for the average coefficient estimates, standard errors, p-values, and coefficient of determination, but for the exit models. These models exhibit similar properties further underscoring the robustness of our approach. The only difference is that some of the coefficients lose statistical significance at certain thresholds.

Finally, in Figure S5 we display histograms of the distribution of the model coefficients when $\Delta t = 10$ and $\tau = 5$. We use this figure to demonstrate that the coefficients are tightly clustered around their mean values, with low variance, indicating consistency across different starting years. This clustering suggests that the model's estimates are similar regardless of the starting year.
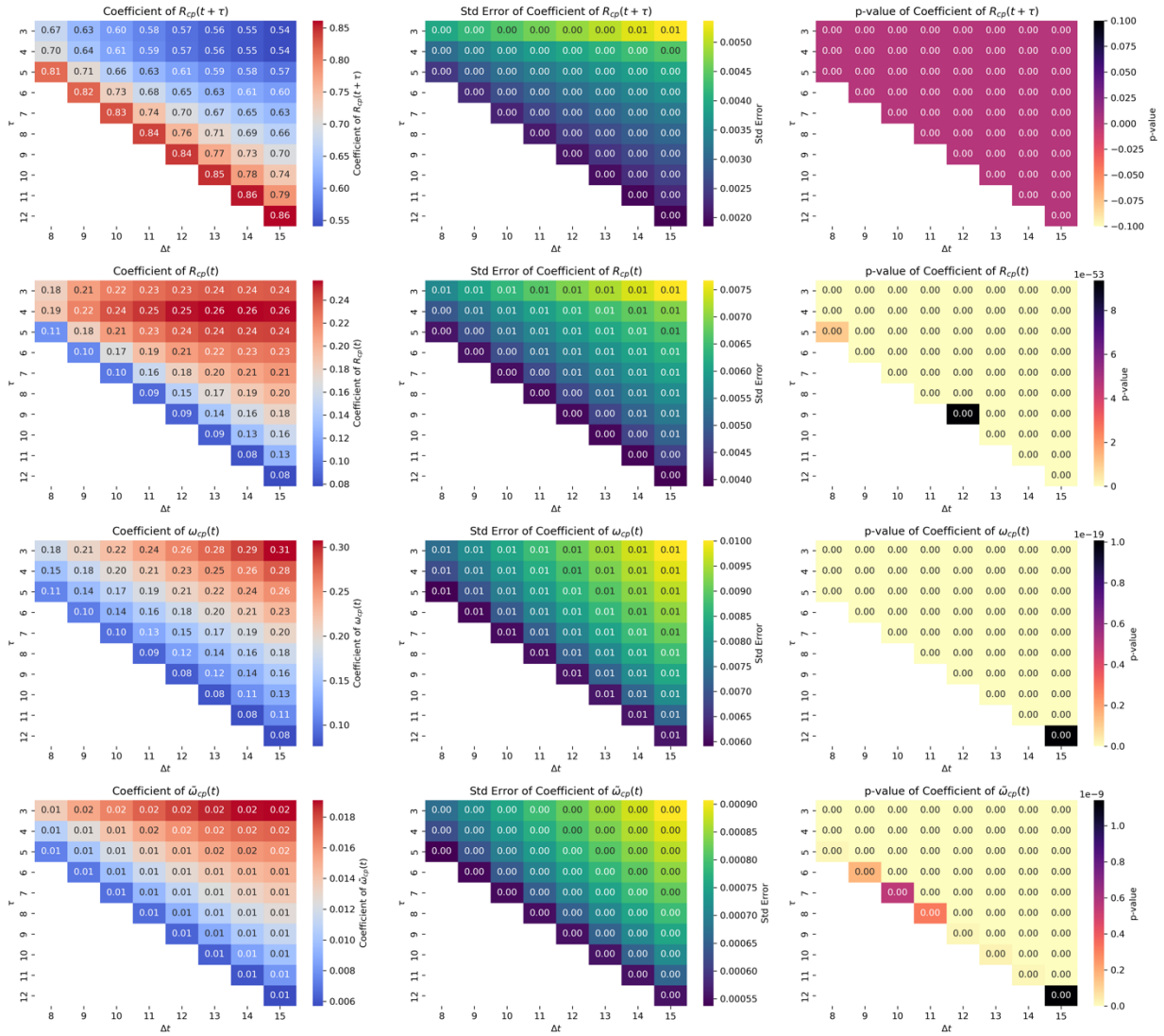
Figure S1. Heatmaps showing the average value of the coefficient estimates, their standard deviation, and their p-values as a function of timeframe Δt and the steppingstone τ, for entry regression models (defined as the subset of the data with $R_{cp}(t) < 1$) estimated using international trade data.
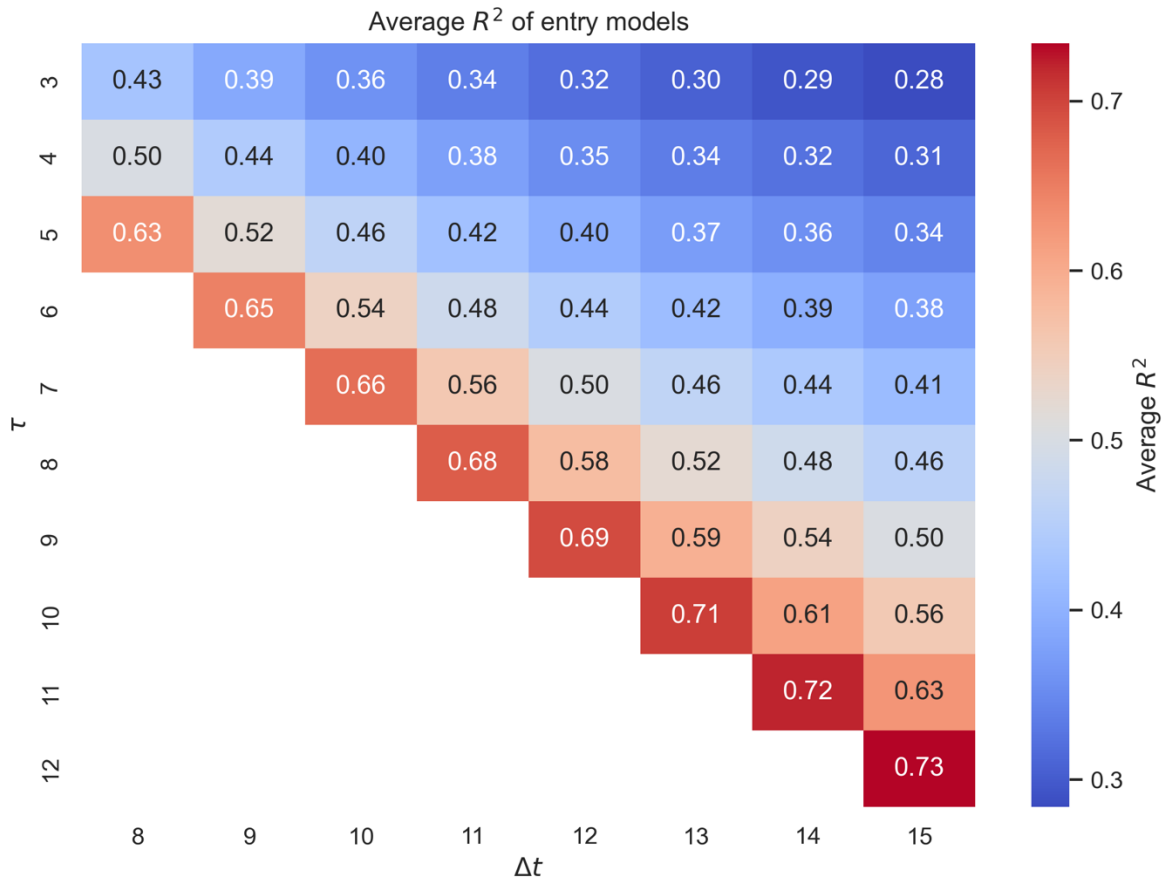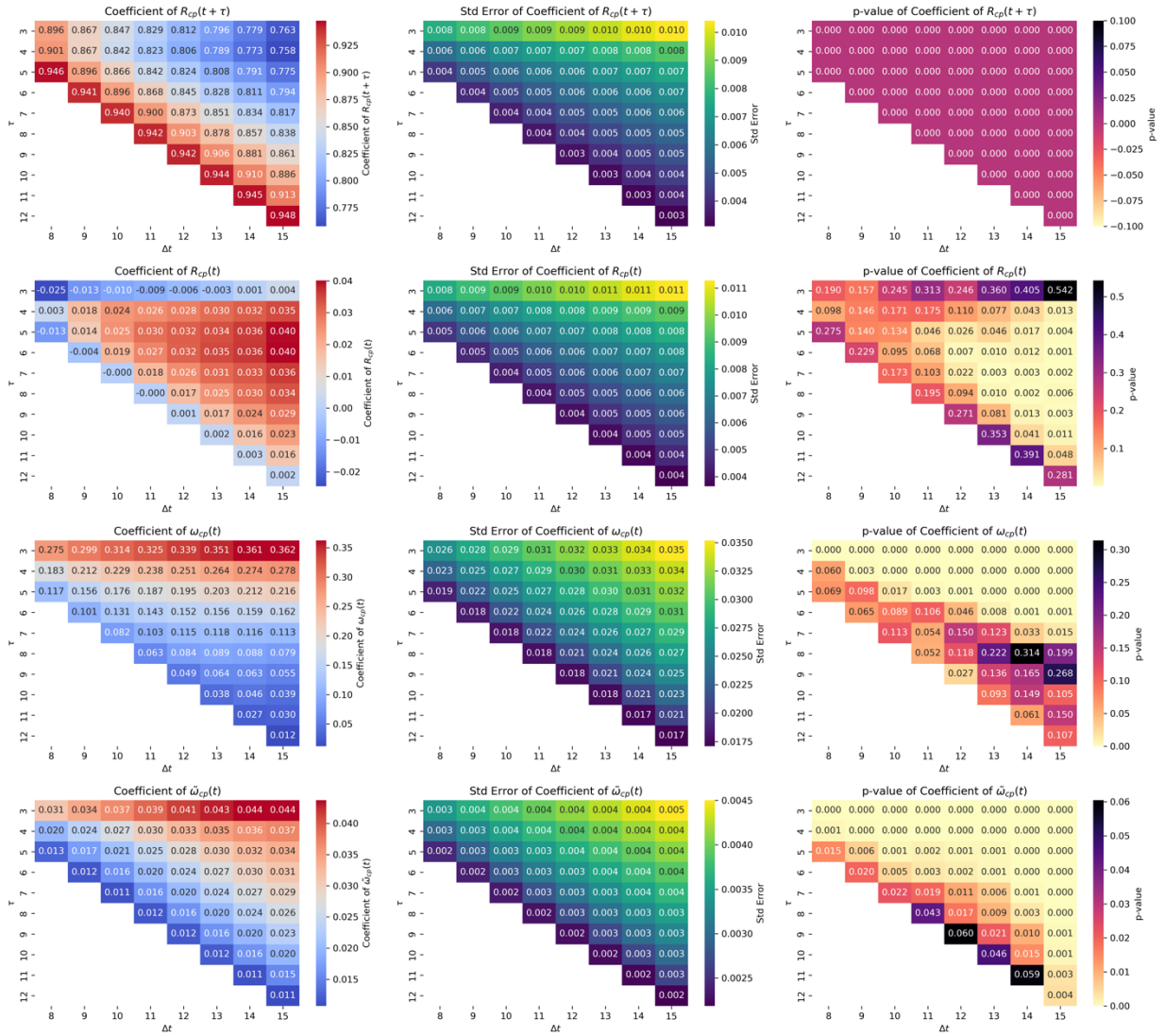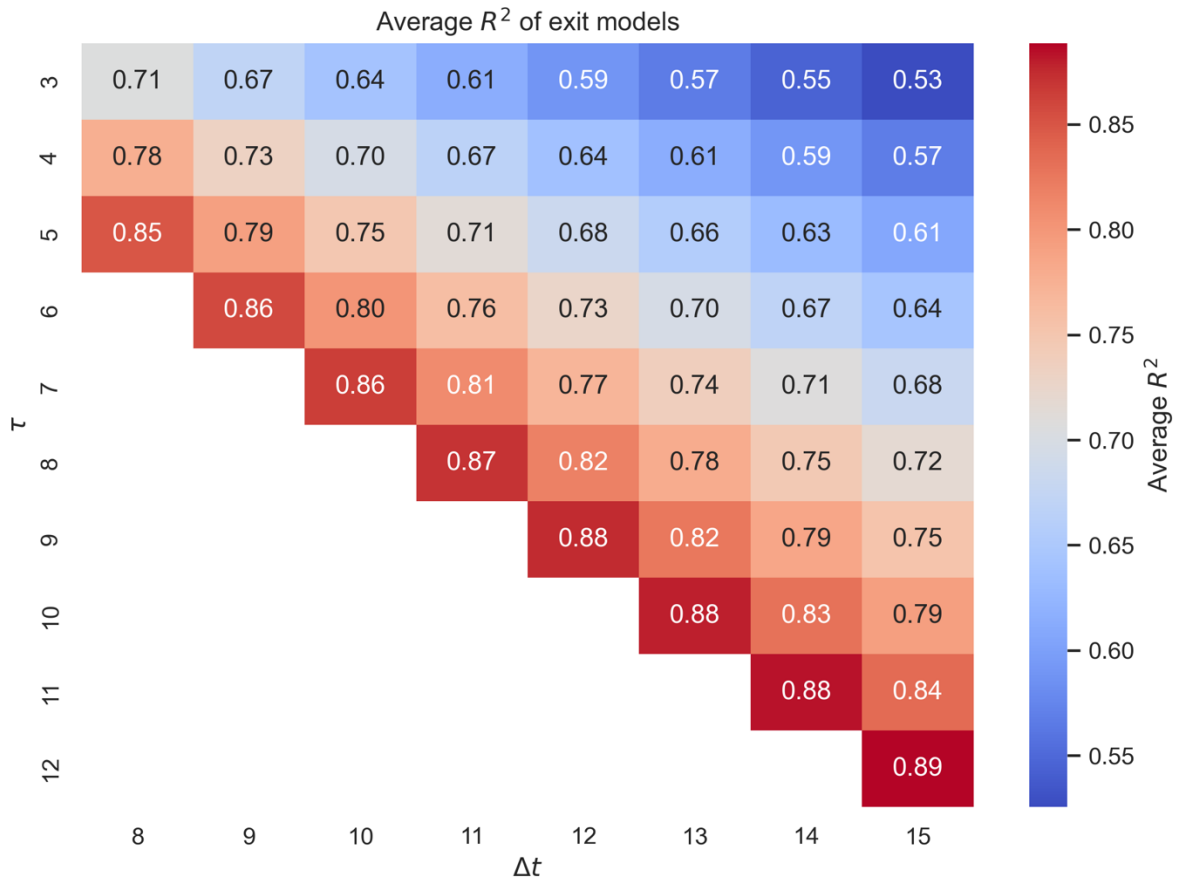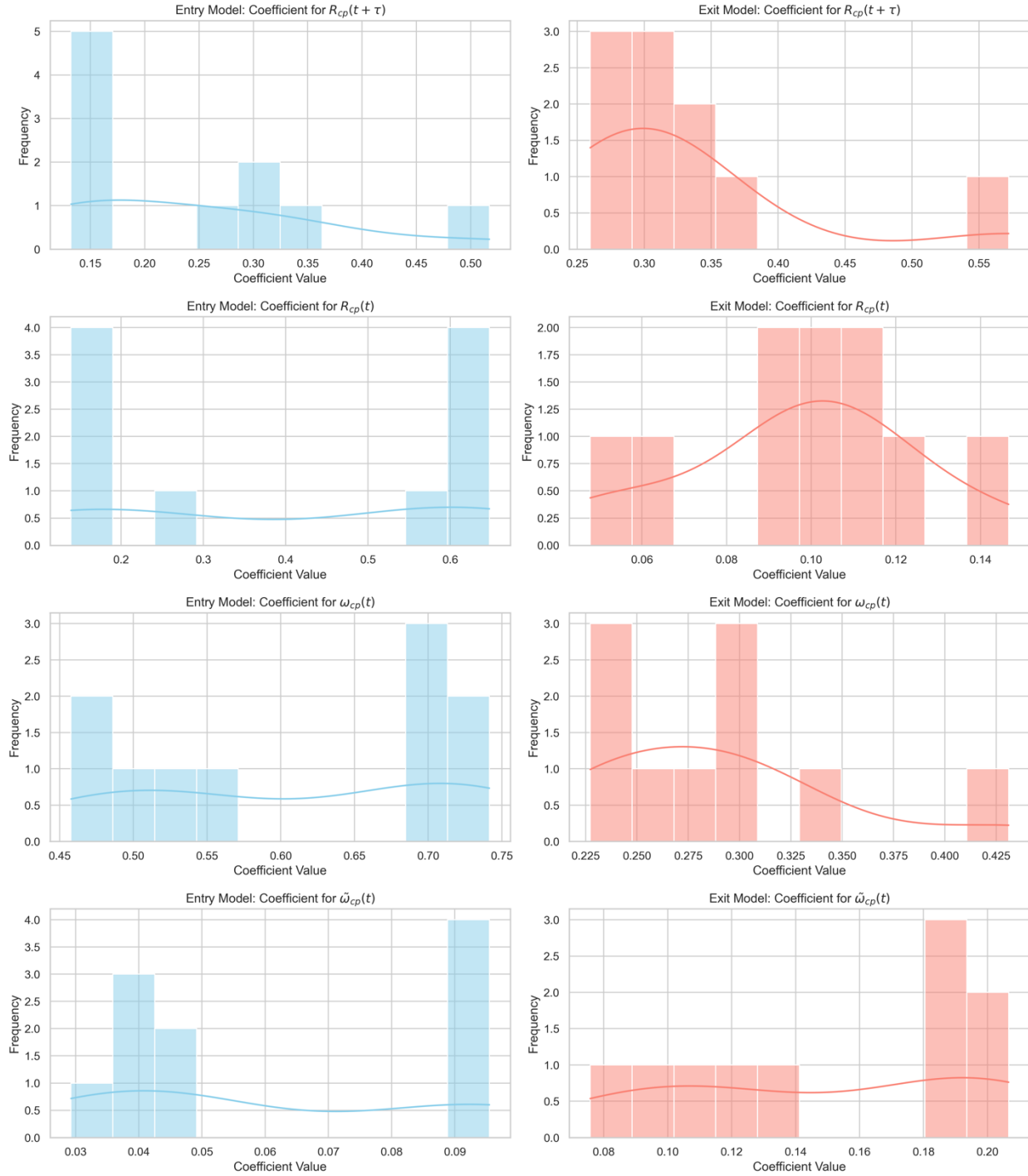
Figure S2. Heatmaps showing the coefficient of determination as a function of timeframe Δt and the steppingstone τ, for entry regression models (defined as the subset of the data with $R_{cp}(t) < 1$) estimated using international trade data.

Figure S3. Heatmaps showing the average value of the coefficient estimates, their standard deviation, and their p-values as a function of timeframe Δt and the steppingstone τ, for exit regression models (defined as the subset of the data with $R_{cp}(t) \geq 1$) estimated using international trade data.

Figure S4. Heatmaps showing the coefficient of determination as a function of timeframe Δt and the steppingstone τ, for exit regression models (defined as the subset of the data with $R_{cp}(t) \geq 1$) estimated using international trade data.

Figure S5. Histogram for the distribution of model coefficients for the models where $\Delta t = 10$ and $\tau = 5$ estimated using international trade data.

Regression Model Performance in The USA MSA Payroll Data

In Figure S6, we present a heatmap of the average coefficient estimates, their standard errors, and p-values for entry models describing equation (2), estimated by using the Usa MSA Payroll data and varying the timeframe Δt and stepping stone τ. Figure S7 complements this by showing the average coefficient of determination for these models. We find that all coefficients exhibit a positive relationship with future specialization, as expected, and that, most of the coefficients are statistically significant at the 0.01 level (only the relatedness coefficient needs a threshold of 0.1 in order to be significant when $\tau = 3$ and $\Delta t = 9$). These models demonstrate an average coefficient of determination of 0.25, indicating a more moderate fit when compared to the international trade data. Nevertheless, the coefficient of determination still follows a gradient pattern, with higher values observed when τ is closer to Δt.

Again, this pattern is primarily due to the autoregressive relationship between the RCA stepping stone $R_{cp}(t + \tau)$ and the dependent variable. As τ approaches Δt, the magnitude of the coefficient for $R_{cp}(t + \tau)$ increases, while the other coefficients correspondingly decrease.

Figures S8 and S9 illustrate the same heatmaps for the average coefficient estimates, standard errors, p-values, and coefficient of determination, but for the exit models. These models exhibit similar properties further underscoring the robustness across the MSA data.

Finally, in Figure S10 we display histograms of the distribution of the model coefficients when $\Delta t = 10$ and $\tau = 5$. We use this figure to demonstrate that the coefficients are tightly clustered around their mean values, with low variance, indicating consistency across different starting years. This clustering suggests that the model's estimates are similar regardless of the starting year.
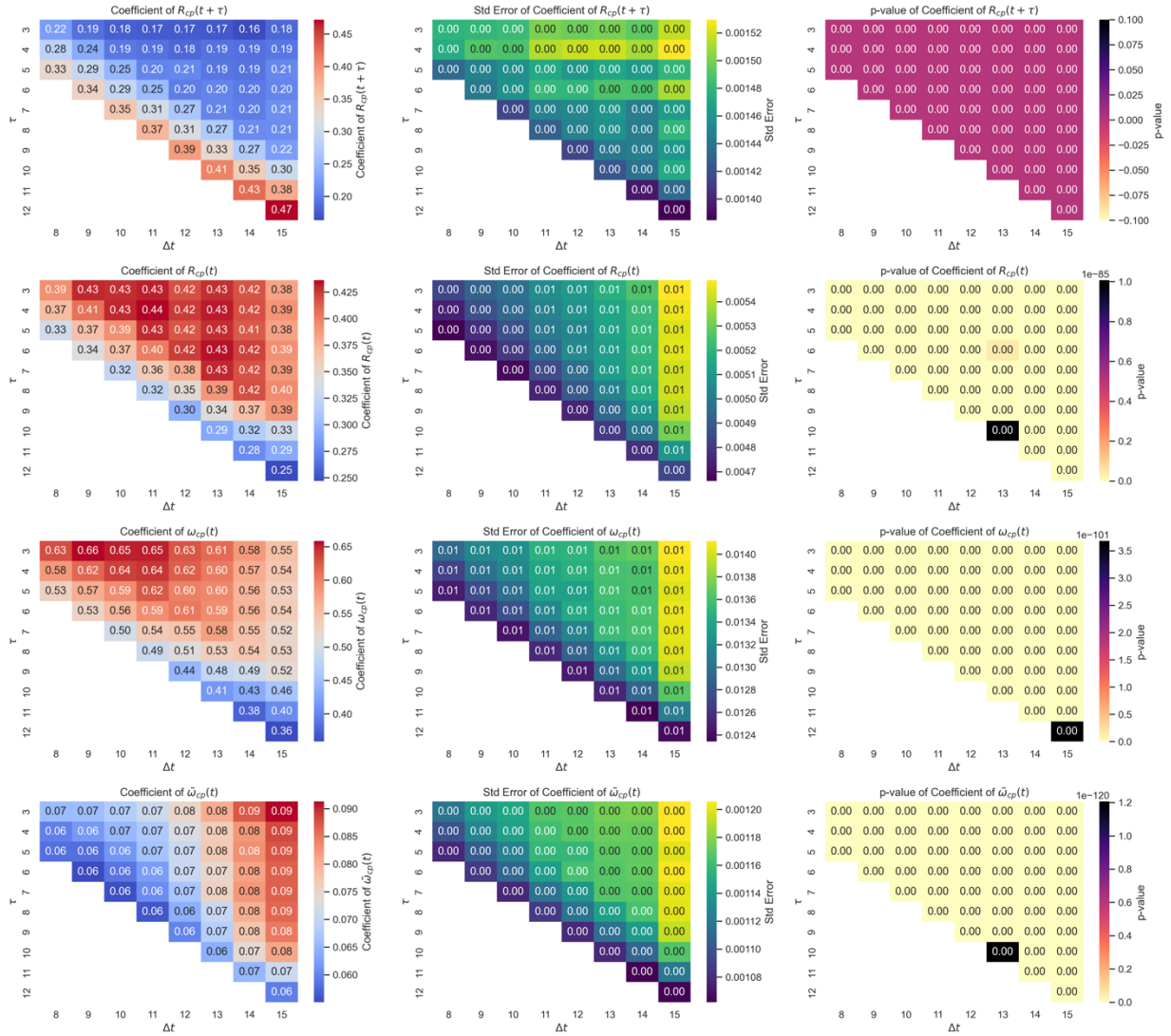
Figure S6. Heatmaps showing the average value of the coefficient estimates, their standard deviation, and their p-values as a function of timeframe Δt and the steppingstone τ, for entry regression models (defined as the subset of the data with $R_{cp}(t) < 1$) estimated using MSA payroll data.
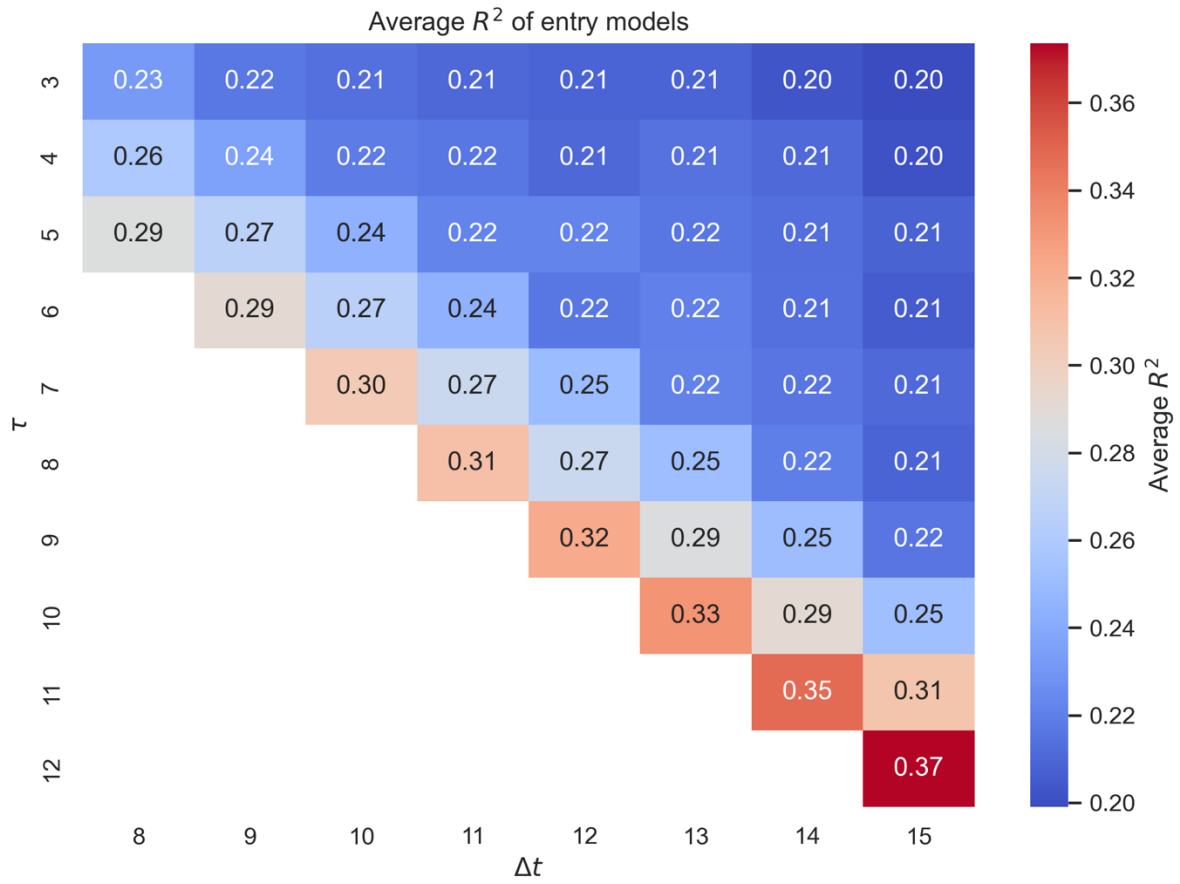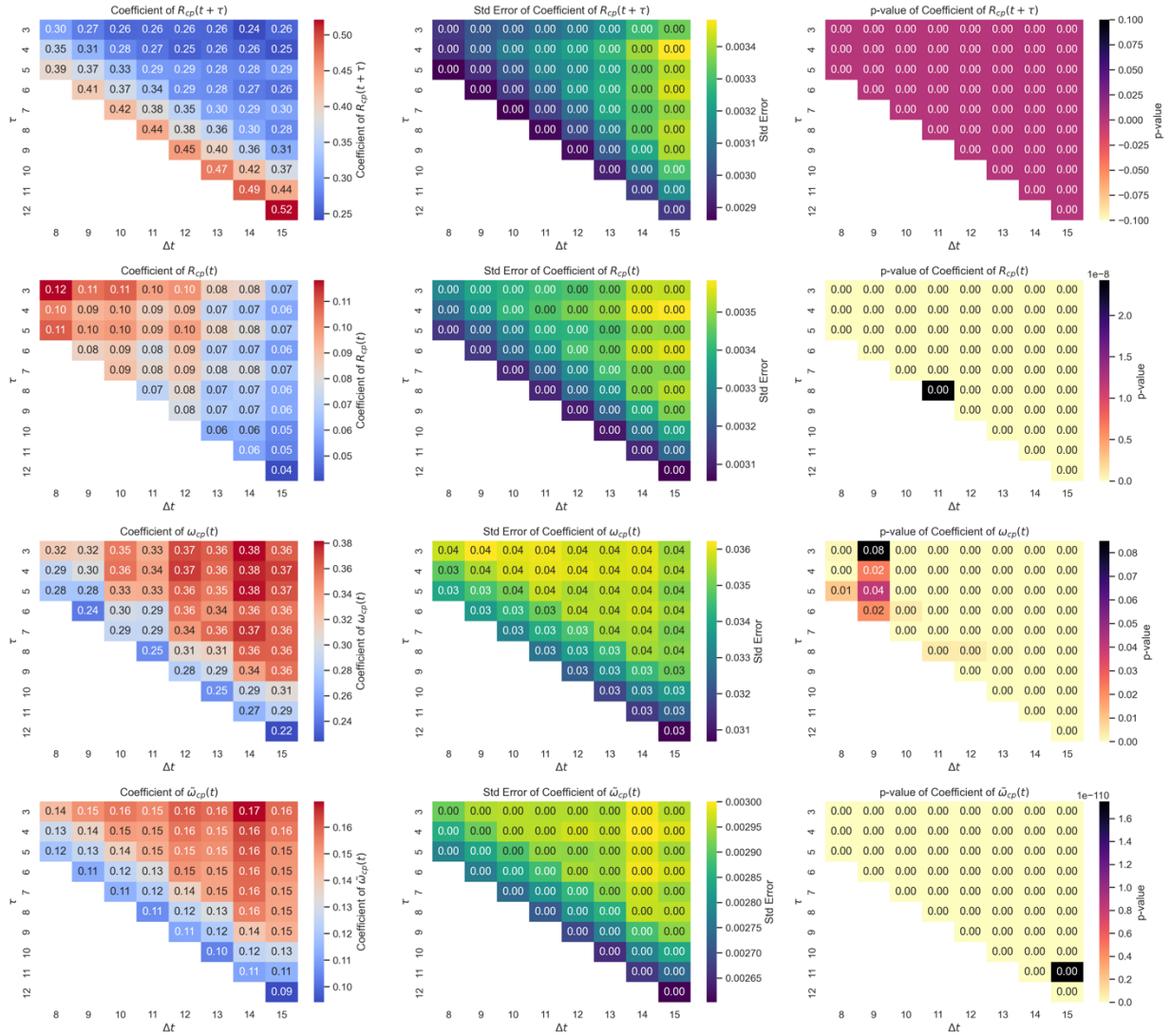
Figure S7. Heatmaps showing the coefficient of determination as a function of timeframe Δt and the steppingstone τ, for entry regression models (defined as the subset of the data with $R_{cp}(t) < 1$) estimated using MSA payroll data.

Figure S8. Heatmaps showing the average value of the coefficient estimates, their standard deviation, and their p-values as a function of timeframe Δt and the steppingstone τ, for exit regression models (defined as the subset of the data with $R_{cp}(t) > 1$) estimated using MSA payroll data.
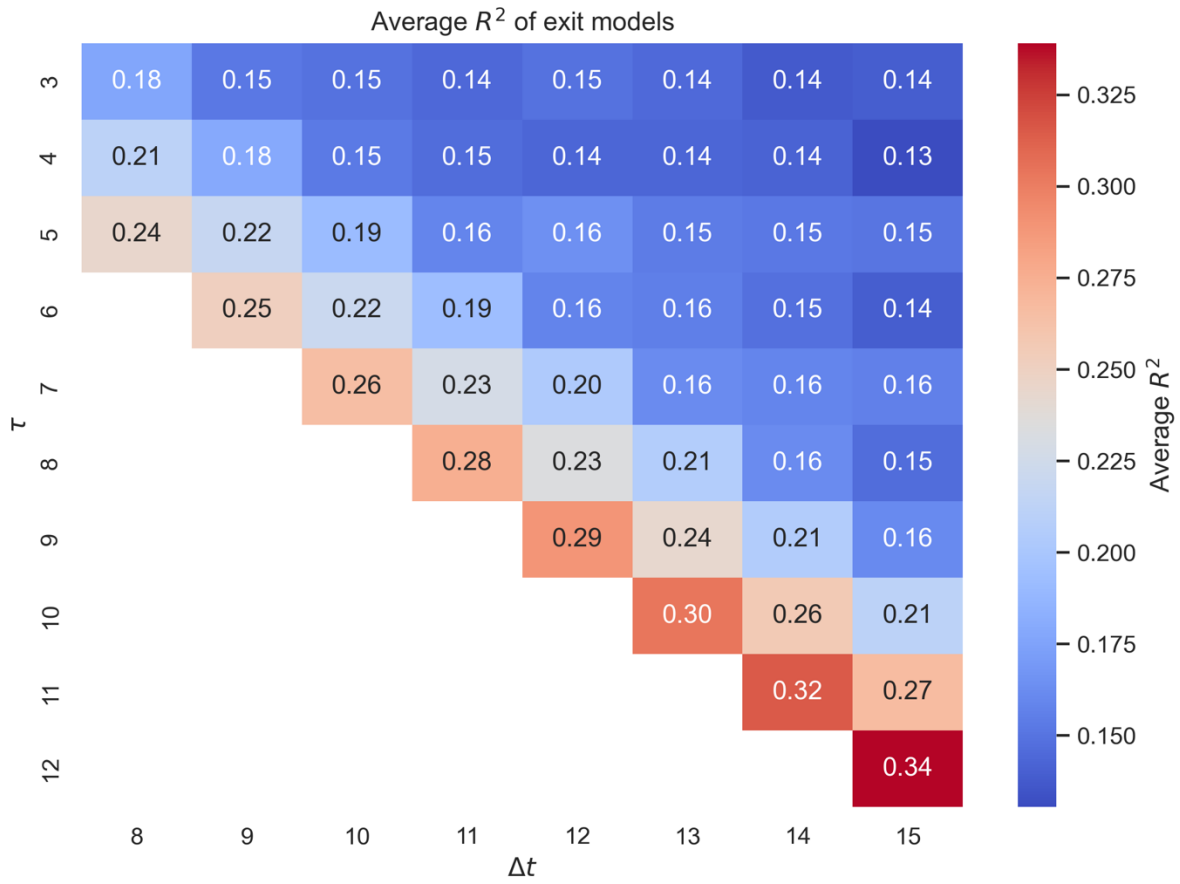
Figure S9. Heatmaps showing the coefficient of determination as a function of timeframe Δt and the steppingstone τ, for exit regression models (defined as the subset of the data with $R_{cp}(t) \geq 1$) estimated using MSA payroll data.
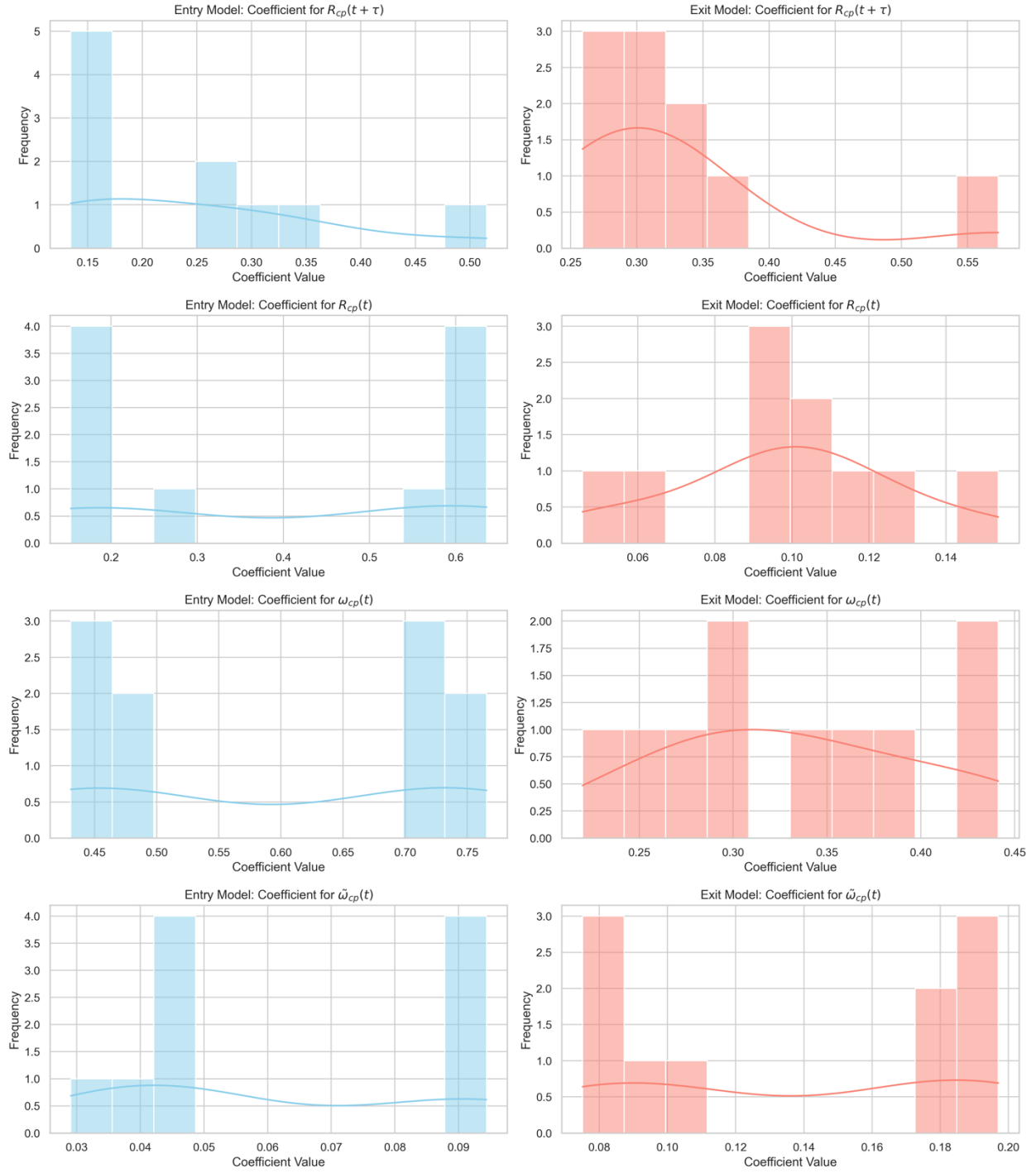
Figure S10. Histogram for the distribution of model coefficients for the models where $\Delta t = 10$ and $\tau = 5$ estimated using MSA payroll data.

Properties of ECI Optimization in the USA MSA Payroll Data

In Figure S11, we repeat this analysis described in the section "Properties of ECI Optimzation" using MSA payroll data. Again, we find that ECI optimization suggests activities that are more aligned with the MSA's current specialization (Figure S11a). But unlike the case of international trade data, we do not observe the U-shaped relationship between ECI and average relative relatedness (Figure S11b). Despite this, ECI optimization still provides suggestions that are of slightly lower relatedness compared to the benchmark model. Additionally, the number of new activities and the added volume suggested by the ECI optimization method are, once again, lower than the benchmark model (Figures S11c and S11d), indicating a balanced approach in terms of both diversification and investment.
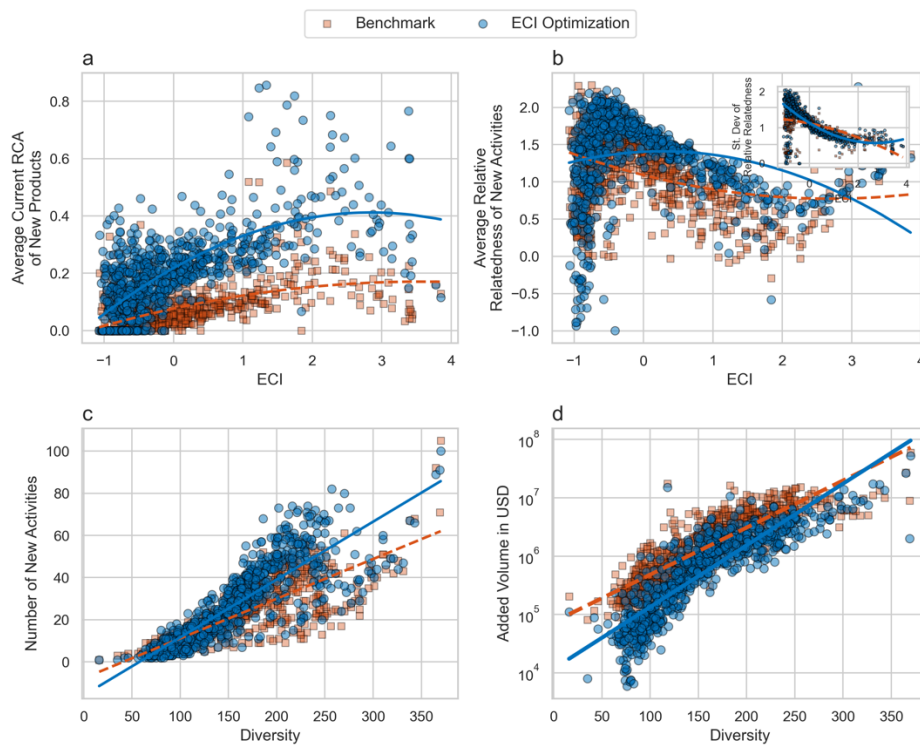

Figure S11. Properties of ECI Optimization in United States MSA Data. **a)** Average RCA (in 2022) of the activities suggested by ECI optimization and the benchmark model as a function of an MSA's ECI in 2022. **b)** Average relative relatedness of the suggested activities as a function of an MSA's ECI in 2022. The inset plot shows the variance of the relative relatedness of the suggested products. **c)** The number of suggested activities as a function of the initial diversity (number of activities in which the MSA had comparative advantage in 2022). **d)** Estimated added

payroll volume to gain comparative advantage in the suggested activities as a function of the initial diversity. **a-d)** For each MSA we assume an increase of 0.1 of the ECI value in 2022. Each scatter chart has a quadratic fit line.

Economic Growth Regression Model Results

In Table S1 we show the results of our economic growth model regressions used in the framework to predict economic structures given a target growth rate.

In these regressions, the dependent variable is the annualized 10-year growth rate of GDP per capita (in PPP constant 2021 USD) of a country (we consider two periods 1999–2009 and 2009–2019). We use two additional explanatory variables. First, we use a z-score normalized value for the log of initial GDP per capita (normalized across our sample of countries for each year), capturing Solow's idea of economic convergence. This z-score transformation helps us account for the non-stationary nature of GDP per capita, and hence provide a consistent prediction. Second, we use the interaction of ECI with the z-score of the initial log of GDP per capita, capturing the idea that the contribution of economic complexity to future economic growth depends on the current level of income.[9] We also use period fixed effects in order to account for any omitted variables that vary over the two decades and may impact economic growth.

In column (1) of Table S1 we show the results of a baseline model incorporating only the Solow term, in column (2) we introduce ECI to the equation, and in model (3) we combine ECI and gdp per capita to predict economic growth. In each case, all of the variables are statistically significant. More importantly ECI significantly upgrades the baselines model predictive power. Namely, the R-squared grows from 0.09 in the baseline model to 0.19 when we use just ECI as a predictor, and to 0.23 when we combine ECI and GDP per capita. This provides a statistical validity for the growth model used in the main manuscript. Nevertheless, we emphasize that more complex models incorporating multiple explanatory variables could potentially increase the explanatory power (but might decrease the available data).

Table S1. Economic Growth Regression Results.

	Dependent variable: Annualized growth of GDP per capita (in PPP constant 2021 USD) (1999-2009, 2009-2019)		
	(1)	(2)	(3)
ECI (trade)		0.868***	1.073***
		(0.199)	(0.199)
ECI (trade) x Log of initial GDP per capita			-0.461***
			(0.150)
Log of initial GDP per capita	-0.714***	-1.392***	-1.536***
	(0.163)	(0.235)	(0.229)
Observations	172	172	172
R^2	0.092	0.192	0.229
Adjusted R^2	0.086	0.182	0.215
Residual Std. Error	1.954 (df=170)	1.848 (df=169)	1.811 (df=168)
F Statistic	19.233*** (df=1; 170)	17.630*** (df=2; 169)	15.409*** (df=3; 168)

Note: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors in parentheses.