HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

Country Diversification, Product Ubiquity, and Economic Divergence

Faculty Research Working Paper Series

Ricardo Hausmann

Harvard Kennedy School

César A. Hidalgo

Harvard Kennedy School

November 2010
RWP10-045

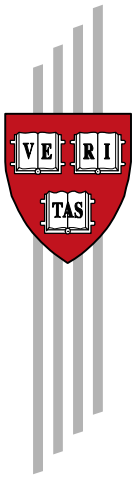
The views expressed in the **HKS Faculty Research Working Paper Series** are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or of Harvard University. Faculty Research Working Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

www.hks.harvard.edu

Country diversification, product ubiquity, and economic divergence

Ricardo Hausmann and César A. Hidalgo
CID Working Paper No. 201
October 2010

© Copyright 2010 Ricardo Hausmann, César A. Hidalgo
and the President and Fellows of Harvard College



Working Papers

Center for International Development
at Harvard University

Country diversification, product ubiquity, and economic divergence

Ricardo Hausmann[1,2] and César A. Hidalgo[1,3]
[1]*Center for International Development, Harvard University*
[2]*Harvard Kennedy School, Harvard University*
[3]*The Media Laboratory, Massachusetts Institute of Technology*

Abstract:

Countries differ markedly in the diversification of their exports. Products differ in the number of countries that export them, which we define as their ubiquity. We document a new stylized fact in the global pattern of exports: there is a systematic relationship between the diversification of a country's exports and the ubiquity of its products. We argue that this fact is not implied by current theories of international trade and show that it is not a trivial consequence of the heterogeneity in the level of diversification of countries or of the heterogeneity in the ubiquity of products. We account for this stylized fact by constructing a simple model that assumes that each product requires a potentially large number of non-tradable inputs, which we call capabilities, and that a country can only make the products for which it has all the requisite capabilities. Products differ in the number and specific nature of the capabilities they require, as countries differ in the number/nature of capabilities they have. Products that require more capabilities will be accessible to fewer countries (i.e., will be less ubiquitous), while countries that have more capabilities will have what is required to make more products (i.e., will be more diversified). Our model implies that the return to the accumulation of new capabilities increases exponentially with the number of capabilities already available in a country. Moreover, we find that the convexity of the increase in diversification associated with the accumulation of a new capability increases when either the total number of capabilities that exist in the world increases or the average complexity of products, defined as the number of capabilities products require, increases. This convexity defines what we term as a *quiescence trap,* or a *trap of economic stasis*: countries with few capabilities will have negligible or no return to the accumulation of more capabilities, while at the same time countries with many capabilities will experience large returns - in terms of increased diversification - to the accumulation of additional capabilities. We calibrate the model to three different sets of empirical data and show that the derived functional forms reproduce the empirically observed distributions of product ubiquity, the relationship between the diversification of countries and the average ubiquity of the products they export, and the distribution of the probability that two products are co-exported. This calibration suggests that the global economy is composed of a relatively large number of capabilities – between 23 and 80, depending on the level of disaggregation of the data – and that products require on average a relatively large fraction of these capabilities in order to be produced. The conclusion of this calibration is that the world exists in a regime where the quiescence trap is strong.

JEL Codes: O11, O14, O33, O57, F43, F47
Keywords: Capabilities, Poverty Trap, Economic Complexity, Structural Transformation, The Product Space, Networks.
Acknowledgments: We would like to acknowledge comments from Pol Antràs, Dany Bahar, Elhanan Helpman, Robert Lawrence, Raja Kali, Lant Pritchett, Roberto Rigobon, Dani Rodrik and Andrés Zahler. We also would like to acknowledge the comments from the audiences at the 2009 International Growth Week Seminar, from the International Growth Center, the CID Faculty Lunch, UNCTAD and the PREM Seminar Series at The World Bank for their comments and questions.

Introduction

In this paper, we develop techniques to characterize the structure of the global pattern of exports and use them to establish and explain the fact that there is a systematic relationship between the number of different products that a country makes, i.e. its diversification, and the number of other countries that on average make those products (which we refer to as the ubiquity of the product). Poorly diversified countries make products that are, on average, made by many other countries, while highly diversified countries make products which are made, on average, by few other countries. This fact is not explained by Ricardian or Heckscher-Ohlin theories of trade and is inconsistent with the basic assumption behind the Dixit-Stiglitz production function (Dixit and Stiglitz 1977) that has become a standard building block of most current trade models. We develop a parsimonious model to account for this fact based on the idea that products are made by combining specific subsets of non-tradable productive inputs, which we will call capabilities. Countries differ in the number and specific combination of the capabilities they have and products differ in the combination of the capabilities they require. We assume that countries only make products for which they have all the required capabilities at their disposal. Because capabilities are by definition non-tradable, their availability determines whether products can be made at a particular location. We derive implications that emerge directly from these assumptions, including the fact that countries with more capabilities will be able to make more kinds of products, while the manufacture of products requiring more capabilities will be accessible to fewer countries.

Moreover, we show that the complementarity of capabilities implies that the increase in diversification that is expected from the accumulation of new capabilities depends strongly on the number of capabilities a country already has. The more capabilities a country has, the higher the return, in terms of increased diversification, that the accumulation of a new capability will provide given that the possible combination of any additional capability with existing ones grows exponentially with the number of capabilities already available in a country. This property of the model creates a *quiescence trap* in the sense that countries with too few capabilities will not have incentives to accumulate additional capabilities, as these are unlikely to be demanded, given the absence of other complementary capabilities. We prove mathematically that this result is independent from any assumption about the initial distribution of capabilities across countries or the distribution of capability requirements across products and show that it stems solely from assuming complementarity in capability requirements. Moreover, we find that there are two ways in which the quiescence trap gets deeper: one is when products are more complex, in the sense that they require a larger fraction of the total number of capabilities, and the other one is when the total number of capabilities in the world becomes relatively large. Both of these alternatives increase the complexity of products in ways that accentuate the quiescence trap, driving the industrial development of different regions of the world towards divergence, rather than convergence.[1]

[1] Since we define capabilities as non-tradable productive inputs, an increase in the tradability of productive inputs -- e.g., through trade in tasks as in Grossman and Rossi-HansbergGrossman, G. A., and E. Rossi-Hansberg, "Trading Tasks: A Simple Theory of Offshoring," American Economic Review, 98 (2008), 1978-1997.-- reduces the quiescence trap: if the value chain can be split up across countries, fewer capabilities have to be present in any

The basic stylized fact that we document and explain is illustrated in Figure 1. Each of these matrices represents how much each country exported of each product. To make countries and products more readily comparable, we control for variations in the size of countries and of product markets by calculating the Revealed Comparative Advantage (RCA) that each country has in each product. For this we use Balassa's (Balassa 1964) definition of RCA as the ratio between the export share of product p in country c and the share of product p in the world market. Formally RCA is defined as:

$$RCA_{cp} = \frac{X_{cp}}{\sum_p X_{cp}} \Big/ \frac{\sum_c X_{cp}}{\sum_{c,p} X_{cp}}, \qquad (1)$$

where X_{cp} represents the dollar exports of country c in product p.

To show that the result is not driven by any particular form of encoding products, we use three different trade classifications systems. The first is the North American Industrial Classification System (NAICS) 6-digit classification. For the year 2006, this dataset contains 132 countries and 318 tradable products categories. The second dataset we use is the Feenstra et al. dataset (Feenstra, Lipsey, Deng, Ma and Mo 2005), which codes products using the SICT4 rev2 classification. We use the most recent year for which this dataset has information, which is the year 2000, containing information for 129 countries and 772 product categories. Finally, we use the Base pour l'Analyse du Commerce International (BACI) dataset from the Centre d'Études Prospectives et d'Informations Internationales (CEPII), which contains data for 232 countries and 5,109 product categories classified using the Harmonized System at the 6-digit level (Gaulier and Zignano 2009).

Figure 1 shows the RCA matrices representing the three datasets described above. In these three examples, rows are sorted according to the *diversification* of countries and columns are sorted by the *ubiquity* of products.[2] We represent the value of the RCA each country has in each product through a color code, which can be read from the figure's color bar.

The matrices of Figure 1 show that in all three datasets, the RCA matrices appear to have a similar, somewhat triangular structure. Some countries (i.e., those described by the first rows of Figure 1), appear to export all products, whereas some products (i.e., those described in the first columns of the figures) appear to be exported by most countries. Moreover, the countries that export few products tend to export the products that almost all countries export, while highly diversified countries export the products that few other countries export. The triangular structure of these matrices suggest that there is a systematic relationship between the diversification of countries and the ubiquity of the products they make: poorly diversified countries have comparative advantage almost exclusively in ubiquitous products, whereas the most diversified countries appear to be the only ones with RCAs in the less ubiquitous products. The fact that the matrix is triangular rather than diagonal suggests that, as countries become more complex, they

---

particular location for the product to be made, making the production of each of the parts being produced at each different location more accessible.

[2] The diversification of countries is calculated simply as the number of products that they export with an RCA above a certain threshold (taken as RCA≥0.5 in this example), whereas the *ubiquity* of products is defined as the number of countries exporting a product with an RCA above a certain threshold (also taken to be RCA≥0.5 in this example). This is done only for the purpose of ordering the matrix. The actual values of RCA are color-coded.

become more diversified; they add more products to the export mix without really abandoning the products they started with.

This fact is not easy to account for using existing trade models. Classical trade theory, whether of the Ricardian or the Heckscher-Ohlin type, tried to explain why countries specialize in different products. As such, these theories take production functions as given and attempt to explain which countries will find it advantageous to specialize in particular sets of products. These theories, however, make no predictions about the number of products made by a country and about the number of countries that make a product. In other words, these theories do not make predictions regarding the diversification of countries, the ubiquity of products, and the relationship between these two dimensions.

New trade theory, on the other hand, (Helpman and Krugmann 1985, Krugman 1979) was developed to account for the increasingly obvious and uncomfortable fact that countries do not appear to specialize. At the basis of that explanation is the assumption that there are scale economies in product development which explain intra-industry specialization. The explanation provided by new trade theory is based on the assumption that products come in varieties and that developing each variety involves some fixed cost. Because these varieties are imperfect substitutes, firms have some market power, but competition erodes their profits so that the monopoly profits they generate in production barely cover the fixed cost of product development. Larger countries have bigger markets in which to amortize the fixed costs of product development and thus would tend to be more diversified. Schott (Schott 2004) and Hummels and Klenow (Hummels and Klenow 2005) provide evidence of this effect.

New trade theory, however, makes no predictions about *which* products will be developed in each country. This is because the theory uses the Dixit-Stiglitz model (Dixit and Stiglitz 1977) which posits a continuum of goods and makes strong assumptions about the symmetry of all goods in order to allow for simple closed-form solutions that are analytically tractable. This eliminates any intrinsic characteristic of the goods considered.

This assumption was originally made for analytical convenience. As argued by Paul Krugman (Krugman 2009):

> "There is *no good reason to believe* that the Dixit-Stiglitz model – a continuum of goods that enter symmetrically into demand, with the same cost functions, and with the elasticity of substitution between any two goods both constant and the same for any pair you choose – *are remotely true in reality*. The assumptions are instead chosen, with full self-consciousness, to produce *a tractable example* that contains what older theories left out – namely, the possibility of intra-industry specialization due to economies of scale."

The Dixit-Stiglitz production function, however, has become a standard building block of most subsequent models of trade, not due to the realism of its assumptions, but in spite of them. There are several elements about the world that get abstracted from view in the DS world. First, the cost of product development is independent of any characteristic of the product, since they are all the same. Second, the cost is also independent of the relationship between a particular

product and the previous productive history of the country. For instance, the cost of developing a regional jet aircraft is the same whether the firm or country has previously developed a transcontinental aircraft a combustion engine or if it produces only raw cocoa and coffee.

Similarly, the Dixit-Stiglitz production function has found its way into theories of growth, where productivity is related to the number of intermediate inputs countries have available for production, with the assumption that the greater the number of intermediate inputs, the higher the productivity with which the economy can operate (Rodriguez-Clare 2007) (Acemoglu, Antras and Helpman 2007). Again, the DS production function assumes that the cost of developing new intermediate inputs is independent of the quantity and nature of the previous intermediate inputs, making the growth process independent of the specific structure of production by assumption while also assuming a link between the number of intermediate inputs and productivity, instead of providing an explanation for why we would expect this to be so.

The Melitz trade model (Melitz 2003), on the other hand, explains which firms would find it advantageous to export and which firms would sell only in the domestic market, but makes no predictions about the number of countries that would have firms that export a particular product or the number of different products exported from a country.

Kremer O-ring model (Kremer 1993) assumes that products differ in the *number* of complementary steps that they require where each step is otherwise identical. In this model, countries with greater ability to perform any step successfully will find it more advantageous to specialize in products that require many steps. Yet, they will be unable to compete with less able countries that specialize in products requiring fewer steps, since wage differentials would make the production of these goods in the most able countries too costly. This model would not predict that high ability countries would be more diversified per se and thus cannot account for the basic stylized fact uncovered in this paper. Indeed, at the limit, the O-ring model predicts that each country produces one category of products (in terms of their difficulty) and each product category is made by one country.

To some extent, our approach is related to the recombinant growth model introduced by Weitzman (Weitzman 1998) or the grammar model introduced by Kauffman (Kauffman 1993). In both, Weitzman's and Kauffman's models the development of new varieties emerges as combinations of previous varieties. Both knowledge of chemistry and optics are required to create photography. In the formalism that we introduce later, this can be interpreted loosely as an increase in the total number of capabilities that exist in the world. Our model differs from that of Weitzman and Kauffman, however, in various dimensions. First, we do not model the historical number of potential varieties that exist in a world, but rather the number of feasible varieties that countries can produce given a limited capability endowment. Second, we use our model to explain differences in the diversity of countries, the ubiquity of products, the connection between these two variables and the probability that a pair of products would be co-exported. Weitzman uses his model to explain the lack of acceleration implied by endogenous growth theory, as an information problem, whereas Kauffman uses his grammar model to explain the historical increase in product diversity, but does not carve implications of his grammar model for the differences in economic diversity and product ubiquity observed in the world. Finally, since the models presented by both Kauffman and Weitzman do not consider connections between sets of

countries and products, they do not make predictions about either the structure of the matrix connecting countries to the products they export, or the Product Space.

Our approach assumes that products are made by combining capabilities and can be represented using binary vectors in which 1's represent the capabilities required by products and 0's represent the capabilities that products do not require. Empirically, we exploit the information contained in the relationship between products and the countries that make them to estimate the relative number of capabilities required by products, and the relative number of capabilities available in countries, following Hidalgo and Hausmann (Hidalgo and Hausmann 2009).

If products are just different combinations of capital and labor, or if they are just arbitrary varieties in some otherwise homogeneous space, then the composition of output should not matter and the mix of products produced or exported by countries should not have serious implications for economic development. Yet, products seem to matter. For Adam Smith, agriculture was bound to be less dynamic than manufactures because it allowed for a more limited division of labor: "The nature of agriculture, indeed does not admit of so many subdivisions of labor, nor of so complete a separation of one business from another, as manufactures" (Smith (1977) [1776]). Here, Smith seems to argue that products differ in a dimension that is not well expressed in any of the standard models of trade. One interpretation is that manufactures require a larger set of capabilities and these capabilities enter, in different combinations, into a larger set of alternative products.

The founders of development economics thought that there was something special about manufactures vis-à-vis agriculture, implying that policies needed to be adopted to achieve structural transformation, an idea that neo-classical economics has had trouble articulating. What makes manufactures so different from agriculture and what do these differences imply, given today's much broader production of goods and services, are questions that have yet to be settled.

In our framework, the obstacles to increased diversification emanate from a fundamental coordination problem: products cannot be made unless the requisite capabilities are present. By the same token, there are no incentives to accumulate any new capability because the products that require them are not being made, so the demand for the capability is initially zero. If there is more than one capability missing for the production of a new product, the provision of any one of the missing capabilities will be of no use. This coordination problem becomes more acute when the number of missing capabilities is larger.

In this framework, products can be considered to be near or far from each other in The Product Space (Hidalgo, Klinger, Barabasi and Hausmann 2007), depending on how many capabilities they share or do not share. Solving the coordination problem is easier for "nearby" products because there are fewer missing capabilities whose provision needs to be coordinated with the demand for them. Economic progress becomes easier if countries make products that have near neighbors in The Product Space, in the sense that they require a similar set of capabilities. Moreover, the ability to develop a new product depends on how many of the requisite capabilities are already present in the country.

Hidalgo et al. (Hidalgo, Klinger, Barabasi and Hausmann 2007) show empirically that products differ in the number of near neighbors they have and that the comparative advantage of countries evolves by moving from the goods they make to those that are nearby in The Product Space. This makes productive transformation more challenging for countries that make products that have few neighbors and these countries grow more slowly on average (Hausmann and Klinger 2006). Also, the presence of nearby products improves the resilience of economies to external shocks. The depth and the duration of recessions triggered by external shocks are correlated with the proximity of current production to alternative products (Hausmann, Rodriguez, and Wagner 2008). Countries that are not well positioned in the Product Space tend to suffer longer and deeper recessions than countries that are better positioned in this network.

Our empirical approach is based on the idea that the availability (requirement) of capabilities in a country (product) can be inferred from export data. Since the presence of a product signals the existence of the requisite capabilities, information on "which country makes what" carries information on which country has what capabilities. Hidalgo and Hausmann (2009) showed that it is possible to count the relative number of capabilities in a country, without making any assumptions on the nature of capabilities, by creating measures that incorporate information that combines the diversification of countries and the ubiquity of products.

The *trick* used to infer the number of capabilities available in a country and required by a product consists in properly mixing information about the diversity of countries and the ubiquity of products. Since we expect countries with more capabilities to produce a wider variety of products – i.e., to be more diversified – than countries with fewer capabilities, diversification is a proxy for the number of capabilities present in a country. This is so because countries with more capabilities will be more likely to have the combinations of capabilities required by more products than countries with fewer capabilities. Hence, the level of diversification of a country will be related to the number of capabilities it has available, albeit imperfectly, since countries producing the same number of products could be making goods that require a different number of capabilities. In such cases, the diversification of countries would not be the most accurate estimator of the number of capabilities available and would need to be corrected by the number of capabilities required by a product. This can be done by looking at the ubiquity of the products made or exported by that country. Products that require few capabilities will be more likely to be produced in many countries and products that require many capabilities will be produced only in the few countries having all the capabilities required. The ubiquity of products, therefore, carries information about their complexity, which can be used to correct diversification as a measure of the number of capabilities available in a country. The incorporation of information on product ubiquity is where our measures depart from other measures of diversification, such as the Herfindahl-Hirschman (Hirschman 1964) index or Entropy (Jost 2006, Saviotti and Frenken 2008), as these other measures do not incorporate any information that differentiates products.

The importance of economic complexity, defined as the relative number of capabilities present in a country and estimated using a procedure in which diversification and ubiquity are used to make sequential corrections for one another, was validated by showing that the estimated number of capabilities in a country correlates strongly with income per capita (R^2~50%) and that deviations from this relationship predict future economic growth (Hidalgo and Hausmann 2009). This suggests that countries approach a level of income which is determined by their capability

endowment, and that these capabilities are expressed, and can therefore be measured, by looking at the mix of products that a country makes.

The remainder of the paper is organized as follow: Section 2 presents the basic stylized fact and analyzes its statistical significance. Section 3 presents the basic model and derives its testable implications. Section 4 calibrates the model to match the different distributions of diversification and ubiquity observed in the data, as well as their correlation. The empirical facts are consistent either with a world of tens of capabilities in which products require a very high proportion of them or with a world of hundreds of capabilities in which products require a small proportion of them. To disentangle these two possibilities, we use the distribution of the probability that products are co-exported. We calibrate the model using three different trade datasets and two different cut-offs and show that, in essence, the world is consistent with the assumption that there are between 23 and 80 capabilities and that products on average require many of them. This implies that we live in a world where the quiescence trap that emerges from the complementarities of the productive value of capabilities is strong. Section 5 concludes.

Section 2. The systematic relationship between diversification and ubiquity

We study the relationship between diversification and ubiquity by introducing a diagram in which the average ubiquity of a country's products is plotted against the number of products exported by that country. We follow the method introduced by Hidalgo and Hausmann (2009) by defining the M_{cp} adjacency matrix, summarizing the connections between countries and the products they export, as:

M_{cp}=1 if country c exports product p with an RCA above a certain R^* threshold,
M_{cp}=0 otherwise.

We calculate the diversification of country c as the sum of M_{cp} over all products

$$k_{c,0} = \sum_p M_{cp}. \quad (2)$$

Similarly, we calculate the ubiquity of product p simply as the sum of M_{cp} over all countries

$$k_{p,0} = \sum_c M_{cp}. \quad (3)$$

In this notation, the average ubiquity of the products exported by country c is defined as:

$$k_{c,1} = \frac{1}{k_{c,0}} \sum_p M_{cp} k_{p,0}, \quad (4)$$

whereas, the average diversification of a product's exporters can be calculated simply as

$$k_{p,1} = \frac{1}{k_{p,0}} \sum_c M_{cp} k_{c,0}. \quad (5)$$

Figure 2 a-f shows the $k_{c,0}$-$k_{c,1}$ (diversification–average ubiquity) diagrams corresponding to the RCA matrices shown in Figure 1 for R^*=1 and R^*=0.5. In all cases, we observe that the average ubiquity of a country's exports tends to decrease with that country's level of diversification. This illustrates that less diversified countries tend to export more ubiquitous products whereas diversified countries are more likely to export products that are also exported by few other countries. A schematic explanation of the $k_{c,0}$-$k_{c,1}$ diagram is presented in Figure 3 a.

Because of the symmetry in the way in which countries and products enter into M_{cp} , it is possible to define an equivalent diagram for products. In the case of products, however, the $k_{p,0}$-$k_{p,1}$ diagram will show the average diversification of the countries' exporting those products as a function of the ubiquity of that product (Figure 3 b).

We test the statistical significance of these patterns by introducing four null models. Since these diagrams summarize structural properties of bipartite networks, their significance can be assessed only by comparing them to bipartite networks with equivalent structural properties

(Figure 3 ç). **Null Model 1** is a random network with the same number of links, that is, with the same average ubiquity and diversification as $\boldsymbol{M}_{cp}$. **Null Model 2** is a randomized network in which the values inside each column of $\boldsymbol{M}_{cp}$ have been shuffled and represent a network in which the diversification of each country matches exactly that observed in the data, yet its exports have been randomly reassigned such that the average ubiquity of the system is conserved. **Null Model 3** is a randomized network in which the values in each row of $\boldsymbol{M}_{cp}$ have been shuffled and represent a network in which the ubiquity of each product matches exactly the one observed in the original data, but the producers of those products have been randomly assigned. The average diversification of Null Model 3 matches that of the original data. **Null Model 4** is a randomized network constructed by permuting the entries of $\boldsymbol{M}_{cp}$ such that the ubiquity of products and diversification of countries remains unchanged. Null Model 4 is the most stringent of the four null models, as it preserves exactly the diversification of each country ($k_{c,0}$) and the ubiquity of each product ($k_{p,0}$) (Figure 3 c). Because of its stringency, however, Null Model 4 does not allow us to randomize the original matrix much in the acute corners of the triangle.

In Figures 3 d and e, we use the $k_{c,0}$-$k_{c,1}$ diagram shown in Figure 2 c (SITC-4 data and R*=1) to illustrate the differences between the structure of M_{cp} and that corresponding to instances of its associated null models. These comparison shows that countries are disproportionately located either higher in the upper-left corner of the $k_{c,0}$-$k_{c,1}$ diagram or deep in its lower-right corner, meaning that the average ubiquity of a country's exports ($k_{p,1}$) decreases with that country's level of diversification ($k_{c,0}$) more abruptly than what would we expect for an ensemble of networks with some of the same structural properties than those observed in the empirical data. The null models also show that the range of variation in diversification and ubiquity observed in the data is much larger than what we would expect from a network with the same number of links. More importantly, null models 2, 3, and 4 show that the negative correlation between the ubiquity of a country's products and its level of diversification cannot be explained simply because some countries export a few products while other countries export many.

We can use the four null models described above to calculate the statistical significance of the slopes observed in the $k_{c,0}$-$k_{c,1}$ and $k_{p,0}$-$k_{p,1}$ diagrams by using them to estimate a p-value for the probability of observing a slope of a certain magnitude in each of these diagrams. Figure 4 illustrates how this procedure was done and summarizes the p-values obtained for the three datasets and two RCA thresholds. The method consists of creating 1000 different instances of the null model, calculating the slopes for each one of them ($S(k_{c,0},k_{c,1})$), and fitting a normal curve to the distribution of slopes obtained from the ensemble of null models. From this fit, it is possible to estimate the probability of observing the slope characterizing each data set given the null model constraints. This test demonstrates that the sharp negative slopes observed in all of the datasets emerges not from the heterogeneity of the distributions of diversification and ubiquity, but rather as a consequence of a non-trivial pattern of connections between countries and products. The case with the lowest statistical significance occurs when we look at Null Model 3, which can be understood quickly by going back to Figure 3. Because Null Model 3 randomizes the diversification of countries while maintaining the ubiquity of products, it creates vertical columns of points in the $k_{c,0}$-$k_{c,1}$ diagram that do not resemble the distribution of points defined by the original data, yet do represent an ensemble of points that is fitted by a wide range of slopes.

While the significance of the negative relationship between $k_{c,0}$-$k_{c,1}$ and $k_{p,0}$-$k_{p,1}$ at R^*=1 and R^*=0.5 is an interesting stylized fact, deviations from the linear relationship contain relevant information. This is because such deviations are informative about the complexity of a country's economy, which can be measured more accurately by looking at the successive averages of these quantities (Hidalgo and Hausmann 2009). An example of this is presented in Figure 3 d, which shows that, while Malaysia and Pakistan export the same number of products the ones exported by Malaysia ($k_{MYS,0}$=104, $k_{MYS,1}$=18) are less ubiquitous than those exported by Pakistan ($k_{PAK,0}$=104, $k_{PAK,1}$=27.5), suggesting that Malaysia's exports are more complex than those of Pakistan, since less ubiquitous products are more likely to require more capabilities. This criterion can be taken further by looking not only at the average ubiquity of Malaysia's and Pakistan's exports, but also by looking at the average level of diversification of the countries that export a similar set of goods as Malaysia or Pakistan. This extended exercise shows that Malaysia exports products that are exported, on average, by countries that are more diversified than the countries that export mixes of goods that are similar to that of Pakistan, suggesting once again that Malaysia export products that require more capabilities than those exported by Pakistan.

The same argument can be used to read the $k_{p,0}$-$k_{p,1}$ diagram. Figure 3 g shows an instance of the $k_{p,0}$-$k_{p,1}$ diagram calculated using the SITC-4 dataset and R*=1, where we have colored products from different sectors according to the ten root categories in the SITC-4 classification. This shows that while there is some correspondence between the $k_{p,0}$-$k_{p,1}$ diagram and the SITC-4 classification, there are important variations among similarly classified products. For example, this graph shows that natural resource-based products such as minerals and fuels exhibit a wide range of ubiquities ($k_{p,0}$), yet are on average exported mostly by not very diversified countries. For instance, coniferous wood, which is highly ubiquitous, is associated with low levels of diversification ($k_{p,0}$=43,$k_{p,1}$=115) as are other less ubiquitous natural resource-based products such as tin ore ($k_{p,0}$=8,$k_{p,1}$=109), suggesting that for this set of products, geography plays a role in reducing their ubiquity by reducing the presence of key natural inputs. On the other hand, products classified as machinery show an important amount of variation in the diversification of their exporters ($k_{p,1}$) at usually relatively low ubiquities ($k_{p,0}$). Hence, the $k_{p,0}$-$k_{p,1}$ diagram can distinguish between simple machines produced in less-diversified countries, such as handheld calculators, ($k_{p,0}$=7,$k_{p,1}$=144) and more complex machines produced in diversified countries, such as motorcycles ($k_{p,0}$=5,$k_{p,1}$=270).

A Simple Model

General Framework

The conspicuous relationships between the diversification of countries and the ubiquity of the products they export motivates us to introduce a simple modeling framework that can be used to understand and reproduce the global patterns of exports summarized in $\boldsymbol{M}_{cp}$: the network connecting countries to the products they export. The model is based on the assumption that production requires the combination of a potentially large number of specific inputs, or capabilities, and that countries can produce goods only if they have all the capabilities that the production of a good requires. In this representation, a country is described as a set of capabilities which can be expressed as a binary vector whose elements are equal to 1, if that country has that capability, and 0 otherwise. In this formalism, products are described by the set of capabilities they require, which can also be expressed as a binary vector in which 1's indicate the capabilities required to produce that product.

Countries can be summarized using a Country-Capability matrix $\boldsymbol{C}_{ca}$, in which each row summarizes the capability endowment of country c, whereas products are specified by the Product-Capability matrix $\boldsymbol{P}_{pa}$, in which each row summarizes the capability requirements of product p. Finally, to specify which countries produce which products we need to define a production function that, given the capability endowment of a country, and the capability requirements of a product, determines whether that country can produce that product or not. Since countries and products are described by matrices summarizing the set of capabilities they have or require, here production is modeled by using an operator that takes $\boldsymbol{C}_{ca}$ and $\boldsymbol{P}_{pa}$ into $\boldsymbol{M}_{cp}$ (it may be helpful to think of the operator as an alternative form of matrix multiplication). Going forward, we assume a world composed by N_c countries, N_p products, and N_a capabilities.

In this interpretation, products require the combination of several inputs, some quite general but others more specific to a smaller set of products. For instance, a shoe manufacturer and a circuit board company both need accountants and a cleaning crew, yet the shoe factory requires workers who are skilled in leather tanning and crusting, as well as leather cutting, sawing, and pasting. The circuit board manufacturing plant, on the other hand, does not need expert leather tanners or seamstresses, but requires people skilled in photo-engraving or PCB milling techniques, which have no use in the shoe factory. Each one of these requirements can be thought of the 1's and 0's which are specified in $\boldsymbol{P}_{pa}$. Yet, in general, we can think that these binary entries include specific infrastructure, regulations, norms, and other non-tradable activities, such as customs and postal services, whose presence or absence can either facilitate or limit the production of these products. Indeed, the formalism we present next helps track the implications of assuming a world in which products require a diverse set of inputs and countries have incomplete sets of inputs, without requiring any definition of what these inputs are, and therefore represent general implications of these fundamental set of assumptions.

Moreover, we assume that each of these products, defined narrowly enough, cannot be produced in the absence of any of the inputs that need to be locally available. This defines C_{ca}. For instance, "tanned leather" cannot be produced without leather tanners and "women shirts" cannot be produced without seamstresses. Hence, we consider that the production of "tanned

leather" by a country strongly suggests the existence of leather tanners in it. This assumption by no means implies no-substitutability. This is because capabilities can be grouped together until a set of purely complementary capabilities is reached and no further substitutions are possible. We assume to be working in that renormalized limit.

The production of products can be thought of as being specified by a Leontief-like production function in which the production of each of these products will be uniquely specified by a combination of inputs and will be equal to zero in absence of any of them. Alternatively, we could think that the production function is some form of a Constant Elasticity of Substitution (CES) but with many potential inputs. If a country lacks any of the inputs that go into a product, output will also be zero. Countries are endowed with some of these inputs, but not others, so that some products are present and others are not. Here we do not concentrate on the intensity with which each of these products is produced, but we focus rather on whether the product is significantly present or not (which we simplify using 0's and 1's).

The discussion below shows how to calculate analytically the predictions for the diversification of countries and the ubiquity of products that emerges from this simple set of assumptions and given forms of C_{ca} and P_{pa}. Our three assumptions are:

A(i) Products require specific combinations of capabilities
A(ii) Countries have some capabilities, but not others
A(iii) Countries will produce goods as long as they have all the required capabilities

The last assumption assumes that countries do not specialize in a subset of the products that are feasible, but rather make them all. This goes against the grain of what much of classical trade theory was about, but the triangular shape of the RCA matrixes suggest that there is little specialization, even when looking at data disaggregated into more than 5,000 product categories; accounting for this lack of specialization is what much of modern trade theory is about.

Formally we define the country-capability and the product-capability matrices as:

$\boldsymbol{C}_{ca}$=1 if country c has capability a and 0 otherwise
$\boldsymbol{P}_{pa}$= 1 if product p requires capability a and 0 otherwise (6)

formalizing assumptions (i) and (ii). In this representation, production is defined as the operator that takes both of these matrices into $\boldsymbol{M}_{cp}$, the matrix connecting countries to products. Formally we denote this as:

$$\boldsymbol{M}_{cp} = \boldsymbol{C}_{ca} \odot \boldsymbol{P}_{pa}. \quad (7)$$

As noted before, $\boldsymbol{M}_{cp}$ is equal to 1 if country c produces product p and 0 otherwise. Assumption (iii) enters in the form of the $\odot$ operator, which based on the assumption presented above, is defined as

$$\boldsymbol{C}_{ca} \odot \boldsymbol{P}_{pa} = 1 \text{ if } \sum_{a}^{N_a} \boldsymbol{C}_{ca}\boldsymbol{P}_{pa} = \sum_{a}^{N_a} \boldsymbol{P}_{pa} \text{ and 0 otherwise.} \quad (8)$$

We refer to this particular form of the ⊙ operator as the Leontief operator, because it resembles a Leontief production function, but in a binary form.

Going forward, we interpret all of these matrices ($\boldsymbol{C}_{ca}$, $\boldsymbol{P}_{pa}$, $\boldsymbol{M}_{cp}$) as bipartite networks connecting countries to the capabilities they have, products to the capabilities they require and countries to the products they make or export. For example, $\boldsymbol{C}_{ca}$=1 is interpreted as a link between country *c* and capability *a*, whereas $\boldsymbol{C}_{ca}$=0 is interpreted as the absence of such a link. $\boldsymbol{M}_{cp}$=1 is interpreted as a link between country *c* and product *p*, meaning that country *c* makes product *p*.

Here we do not adopt any a priori definition of capabilities and therefore consider $\boldsymbol{C}_{ca}$ and $\boldsymbol{P}_{pa}$ as empirically unobservable quantities. $\boldsymbol{M}_{cp}$ is therefore the main prediction of the model and we will compare its structure with that of empirical data through four different observables.

- O(i) The relationship between a country's diversification and the average ubiquity of its products.
- O(ii) The relationship between the ubiquity of a product and the average level of diversification of the countries exporting it.
- O(iii) The distribution of diversification: the probability that a country exports a given number of products.
- O(iv) The distribution of ubiquity: the probability that a product is exported by a given number of countries.

To differentiate between the number of links connecting a country to the products it makes from the number of links connecting a country to the capabilities it has, we use the superscripts (*p*) for products and (*a*) for capabilities. Hence, we define,

$$k_{c,0}^{p} = \sum_{p=1}^{N_p} \boldsymbol{M}_{cp} \tag{9}$$

$$k_{p,0}^{c} = \sum_{c=1}^{N_c} \boldsymbol{M}_{cp}. \tag{10}$$

$$k_{c,1}^{p} = \frac{1}{k_{c,0}^{p}} \sum_{p=1}^{N_p} \boldsymbol{M}_{cp} k_{p,0}^{c} \tag{11}$$

$$k_{p,1}^{c} = \frac{1}{k_{p,0}^{c}} \sum_{c=1}^{N_c} \boldsymbol{M}_{cp} k_{c,0}^{p} \tag{12}$$

$$k_{c,0}^{a} = \sum_{a=1}^{N_a} \boldsymbol{C}_{ca} \tag{13}$$

$$k_{p,0}^{a} = \sum_{a=1}^{N_a} \boldsymbol{P}_{pa} \tag{14}$$

In the next section, we show how to calculate the predictions for the observables O(i)-O(iv) that emerge from assumptions A(i)-A(iii) using a particular case of the combinatorial model in which $\boldsymbol{C}_{ca}$ and $\boldsymbol{P}_{pa}$ are fully random matrices.

The binomial model

Here we study the case in which $\boldsymbol{C}_{ca}$=1 with probability r and 0 with probability 1-r, and $\boldsymbol{P}_{pa}$=1 with probability q and 0 with probability 1-q.

First we use this model to calculate how the diversification of countries ($k_{c,0}^{p}$), and the ubiquity of products ($k_{p,0}^{c}$), depends on the model parameters (N_a, r and q) and on the number of capabilities that a country has ($k_{c,0}^{a}$) and a product requires ($k_{p,0}^{a}$). We build on these results to show that under these assumptions (A(i)-A(iii)):

i) The level of diversification of a country increases on average with the number of capabilities it has:
$dk_{c,0}^{p}/dk_{c,0}^{a}>0$

ii) The ubiquity of a product decreases on average with the number of capabilities it requires.
$dk_{p,0}^{c}/dk_{p,0}^{a}<0$

iii) The average ubiquity of products exported by a country decreases with that country's level of diversification.
$dk_{c,1}^{p}/dk_{c,0}^{p}<0$

iv) The average level of diversification of a product's exporters decreases with the ubiquity of that product.
$dk_{p,1}^{c}/dk_{p,0}^{c}<0$

After this, we calculate the probability that a country exports $k_{c,0}^{p}$ products and that a product is exported by $k_{p,0}^{c}$ countries. We will interpret these results in the following section and we will calibrate them to the real world in the section after that.

Proving (i) is equivalent to finding a function that would relate the average level of diversification of a country to the number of capabilities it has. Namely we are searching for

$$\overline{k_{c,0}^{p}}\left(k_{c,0}^{a}\right). \qquad (15)$$

Where overhead bars are used to denote averages, or expected values. For instance, in equation (15) $\overline{k_{c,0}^{p}}(k_{c,0}^{a})$ represents the average diversification of a country with $k_{c,0}^{a}$ capabilities.

We calculate the average number of products produced by a country with $k_{c,0}^{a}$ capabilities by adding over all products that require a given number of capabilities: the probability that that country makes a product requiring that number of capabilities, times, the number of products that require that number of capabilities. Mathematically we represent this in a general form as:

$$\overline{k_{c,0}^{p}}=\sum_{x=0}^{N_a}\pi\left(c\left(k_{c,0}^{a}\right)\rightarrow p\left(k_{p,0}^{a}=x\right)\right)N_p(k_{p,0}^{a}=x), \qquad (16)$$

where $\pi\left(c\left(k_{c,0}^{a}\right) \rightarrow p\left(k_{p,0}^{a} = x\right)\right)$ represents the probability that a country *c*, with $k_{c,0}^{a}$ capabilities exports a product *p* requiring $k_{p,0}^{a} = x$ capabilities. The number of products requiring *x* capabilities is represented by $N_p(k_{p,0}^{a} = x)$.

We calculate $\overline{k_{c,0}^{p}}$ for a particular country by considering the number of capabilities that the country has as given and equal to: $k_{c,0}^{a}$. This implies that the realization of the random variable *r* for that country is equal to the number of capabilities it has over the number of capabilities that exist, or $(\frac{k_{c,0}^{a}}{N_a})$. Additionally, from the Leontief operator, the probability that a country with $k_{c,0}^{a}$ capabilities exports a product requiring $k_{p,0}^{a} = x$ capabilities, is given by the probability that that country has all the capabilities required by that product. As in this model, the capabilities that a country has are independent random variables, the probability that a country has the *x* capabilities required to produce a product is given by:

$$\pi\left(c\left(k_{c,0}^{a}\right) \rightarrow p\left(k_{p,0}^{a} = x\right)\right) = \left(\frac{k_{c,0}^{a}}{N_a}\right)^{x}. \tag{17}$$

Since products require a capability with a probability *p*, the probability that a product requires *x* capabilities is given by a binomial distribution (which is why we call this implementation the binomial model). Hence,

$$N\left(k_{p,0}^{a} = x\right) = N_p \binom{N_a}{x} q^{x}(1-q)^{N_a - x}. \tag{18}$$

Using (17) and (18), we can take (16) into the model specific form

$$\overline{k_{c,0}^{p}} = N_p \sum_{x=0}^{N_a} \left(\frac{k_{c,0}^{a}}{N_a}\right)^{x} \binom{N_a}{x} q^{x}(1-q)^{N_a - x}, \tag{19}$$

which can be simplified using the binomial theorem, or Newton's Binomial, to

$$\overline{k_{c,0}^{p}} = N_p \left(q\,\frac{k_{c,0}^{a}}{N_a} + 1 - q\right)^{N_a}. \tag{20}$$

It is trivial to show from (20) that the expected number of products produced by a country is a monotonically increasing function of the number of capabilities it has[3].

$$\frac{d\overline{k_{c,0}^{p}}}{dk_{c,0}^{a}} = qN_p \left(q\,\frac{k_{c,0}^{a}}{N_a} + 1 - q\right)^{N_a - 1} \geq 0. \tag{21}$$

In eqn. (21), the equality holds only for the case in which a country has all capabilities ($k_{c,0}^{a} = N_a$) and products require all capabilities (q=1). This concludes our demonstration of (i).

Next, we calculate the expected ubiquity of a product requiring $k_{p,0}^{a} = x$ capabilities. In a model independent form, this can be expressed as:

$$\overline{k_{p,0}^{c}} = \sum_{x=0}^{N_a} \pi\left(c(k_{c,0}^{a} = x) \rightarrow p\left(k_{p,0}^{a}\right)\right) N_c(k_{c,0}^{a} = x), \tag{22}$$

[3] Notice that for large N_a, equation (20) reduces to the exponential form: $\overline{k_{c,0}^{p}} = N_p \exp(q(k_{c,0}^{a} - N_a))$

where $N_c\left(k_{c,0}^a = x\right)$ is the number of countries that have x capabilities and $\pi\left(c(k_{c,0}^a = x) \rightarrow p\left(k_{p,0}^a\right)\right)$ is the probability that a country with *x* capabilities makes a product that requires $k_{p,0}^a$ capabilities. We take (22) into this binomial model by considering $k_{p,0}^a$ as fixed for the product under study, and that the probability for a randomly chosen country to export a product requiring $k_{p,0}^a$ capabilities is given by the probability that it has each of the $k_{p,0}^a$ capabilities that that product requires.

$$\pi\left(c\left(k_{c,0}^a = x\right) \rightarrow p\left(k_{p,0}^a\right)\right) = \left(\frac{x}{N_a}\right)^{k_{p,0}^a}. \tag{23}$$

Which we approximate as

$$\pi\left(c\left(k_{c,0}^a = x\right) \rightarrow p\left(k_{p,0}^a\right)\right) = r^{k_{p,0}^a}. \tag{24}$$

Finally, we consider that $N\left(k_{c,0}^a = x\right)$ is given by N_c times a binomial distribution $N\left(k_{c,0}^a = x\right) \sim \mathrm{N_c B(N_a}, r)$ and that $r^{k_{p,0}^a}$ comes out of the sum in (22) as it does not depend on the summand *x*. Hence, in this particular case, eqn. (22) simplifies to the expression

$$\overline{k_{p,0}^c} = N_c r^{k_{p,0}^a}. \tag{25}$$

From (25) it is trivial to show that the ubiquity of a product is a decreasing function of the number of capabilities it requires, concluding our demonstration of (ii).

Using our previously introduced notation, we can calculate the average ubiquity of the products exported by a country in a general form as:

$$\overline{k_{p,1}^c} = \frac{\sum_x^{N_a} \pi\left(c\left(k_{c,0}^a\right) \rightarrow p\left(k_{p,0}^a = x\right)\right) N_p\left(k_{p,0}^a = x\right) k_{p,0}^c\left(k_{p,0}^a = x\right)}{\sum_x^{N_a} \pi\left(c\left(k_{c,0}^a\right) \rightarrow p\left(k_{p,0}^a = x\right)\right) N_p\left(k_{p,0}^a = x\right)}. \tag{26}$$

Which can be taken into a model specific form as:

$$\overline{k_{p,1}^c} = \frac{\sum_x^{N_a} \left(\frac{k_{c,0}^a}{N_a}\right)^x \binom{N_a}{x} q^x (1-q)^{N_a - x} N_c r^x}{\sum_x^{N_a} \left(\frac{k_{c,0}^a}{N_a}\right)^x \binom{N_a}{x} q^x (1-q)^{N_a - x}}. \tag{27}$$

Using the binomial theorem this simplifies to

$$\overline{k_{c,1}^p} = \frac{N_c \left(rq \frac{k_{c,0}^a}{N_a} + 1 - q\right)^{N_a}}{N_p \left(q \frac{k_{c,0}^a}{N_a} + 1 - q\right)^{N_a}}. \tag{28}$$

To obtain $k_{c,0}^a(\overline{k_{c,0}^p})$ we invert (20) and insert it into (28). After some algebra, we obtain an expression for the average ubiquity of a country's products as a function of its diversification, obtaining our prediction for the observable O(i).

$$\overline{k_{c,1}^p} = \frac{N_p N_c}{\overline{k_{c,0}^p}} \left(r \left(\frac{\overline{k_{c,0}^p}}{N_p} \right)^{1/N_a} + (1-q)(1-r) \right)^{N_a}. \tag{29}$$

Finally, we differentiate (29) with respect to $k_{c,0}^p$ to obtain

$$\frac{d\overline{k_{c,1}^{p}}}{d\overline{k_{c,0}^{p}}} = -\frac{N_p N_c}{\left(\overline{k_{c,0}^{p}}\right)^2}(1-q)(1-r)\left(r\left(\frac{\overline{k_{c,0}^{p}}}{N_p}\right)^{\frac{1}{N_a}} + (1-q)(1-r)\right)^{N_a-1}, \quad (30)$$

which is negative as long as q<1 and r<1, proving that in this model the structure of $\boldsymbol{M}_{cp}$ is such that the ubiquity of a country's products decreases with that country's level of diversification, concluding our demonstration of (iii).

Finally we use the binomial model to calculate the relationship between the average level of diversification of a product's exporters ($\overline{k_{p,1}^{c}}$) and the average ubiquity of a product ($\overline{k_{p,o}^{c}}$) to obtain our prediction for observable O(ii):

$$\overline{k_{p,1}^{c}} = \frac{N_p}{\overline{k_{p,o}^{c}}}\sum_{x=0}^{N_a}\binom{N_a}{x} r^x (1-r)^{N_a-x}\left(\frac{x}{N_a}\right)^{\log \overline{k_{p,o}^{c}}/\log(r)}\left(\frac{qx}{N_a}+1-q\right)^{N_a}. \quad (31)$$

While numerical evidence indicates that, according to (31), the average diversification of a country's exporters will tend to decrease with the ubiquity of a product, we have yet no analytical proof of this.

Finally we use the binomial model to calculate the probability that a country has a given level of diversification or that a product has a given ubiquity. We do this by using a mathematical identity that connects two random variables, x and y, that are related by the function $x=g(y)$. The identity states that the distribution $P(y)$ followed by the random variable y, will be related to the distribution $f(x)$ followed by the random variable as:

$$P(y) \sim f\big(g(y)\big)\frac{dg(y)}{dy}. \quad (32)$$

To calculate the diversification distribution that emerges from the binomial model, we first invert (20) to find $k_{c,0}^{a}$ as a function of $k_{c,0}^{p}$:

$$k_{c,0}^{a}(k_{c,0}^{p}) = \frac{N_a}{q}\left(\left(\frac{k_{c,0}^{p}}{N_p}\right)^{1/N_a} + q - 1\right) \quad (33)$$

and then find the rate of change of $k_{c,0}^{a}$ as a function of $k_{c,0}^{p}$.

$$\frac{dk_{c,0}^{a}}{dk_{c,0}^{p}} = \frac{1}{qN_p}\left(\frac{k_{c,0}^{p}}{N_p}\right)^{(1-N_a)/N_a} \quad (34)$$

Using (33) and (34), and the fact that the number of capabilities in a country follows a binomial distribution, we can show that the distribution of diversifications that emerges from the binomial model is given by

$$P\big(k_{c,0}^{p}\big) \sim \frac{1}{q}\left(\frac{k_{c,0}^{p}}{N_p}\right)^{\frac{1-N_a}{N_a}}\left(\frac{N_a}{q}\left(\left(\frac{k_{c,0}^{p}}{N_p}\right)^{1/N_a}+q-1\right)\right)^{N_a} r^{\frac{N_a}{q}\left(\left(\frac{k_{c,0}^{p}}{N_p}\right)^{1/N_a}+q-1\right)}(1-r)^{N_a-\frac{N_a}{q}\left(\left(\frac{k_{c,0}^{p}}{N_p}\right)^{1/N_a}+q-1\right)} \quad (35)$$

Which can be simplified by considering $N_a \gg 1$ and a change of variables that looks at the fraction of the total number of products rather than its absolute number ($u = k_{c,0}^{p}/N_p$)

$$P(u) \sim \frac{N_p}{qu}\left(\begin{matrix} N_a \\ \frac{N_a}{q}\left(u^{1/N_a}+q-1\right) \end{matrix}\right) r^{\frac{N_a}{q}\left(u^{1/N_a}+q-1\right)}(1-r)^{N_a-\frac{N_a}{q}\left(u^{1/N_a}+q-1\right)} \qquad (36)$$

Similarly, we can show that in this model the distribution of the fraction of countries that export a product ($v = k_{p,0}^{c}/N_c$) is given by:

$$P(v) \sim \frac{N_c}{v\log(r)}\binom{N_a}{\log(v)/\log(r)} q^{\log(v)/\log(r)}(1-q)^{N_a-\log(v)/\log(r)} \qquad (37)$$

This concludes our derivation of observables O(iii) and O(iv) (eqns. (36) and (37)).

Implications of the Binomial Model

In the previous sections we introduced a framework that can be used to model the structure of the network connecting countries to the products they export and we implemented it for a particular case in which the probability that a country (product) has (requires) a given number of capabilities, is both constant and equal for all countries (products). We modeled the network connecting countries to the products they export, $\boldsymbol{M}_{cp}$, as a combination of both of these matrices by defining an operator that connects a country to a product if and only if that country has all the capabilities that the product requires. Moreover, we showed that the country-product matrix ($\boldsymbol{M}_{cp}$) that results from combining these two random matrices is not random, since the $\odot$ operator introduces correlations between the level of diversification of countries and the ubiquity of the products they make that are captured in expressions (29) and (31). Assumption A (iii), which states that countries can make products if and only if they have all the required capabilities, implies that the ubiquity of a country's products decreases with the level of diversification of a country, a prediction that corresponds qualitatively with our empirical findings.

The binomial model also makes some purely theoretical predictions about the world. Equation (20) shows the prediction of the binomial model for the relationship between the number of products that a country makes and the number of capabilities it has. From either (20) or (21) it is trivial to show that $\frac{d^2 k_{c,0}^{p}}{dk_{c,0}^{a\ 2}} > 0$ for N_a>1. This indicates that the marginal effect of adding a capability on the diversification of a country, seen as the expected increase in the number of products that result from an additional capability, increases with the number of capabilities that a country has. In other words, in this model, the accumulation of capabilities expresses itself in a convex increase in diversification for N_a>1. In this model, however, the "increasing returns" in diversification that are expressed through this convexity are not an assumption of the model, but an outcome of it. Moreover, this result is valid for all Leontief type operators and is not affected by the shape of the $\boldsymbol{C}_{ca}$ and $\boldsymbol{P}_{pa}$ matrices.

Indeed, we can prove that, in general, diversification is a convex function of the number of capabilities a country has by taking the model independent form for $\overline{k_{c,0}^{p}}(k_{c,0}^{a})$ (eqn. (16)). We can construct this proof simply by noticing that the number of products that require a given number of capabilities $\left(N_p\left(k_{p,0}^{a}=x\right)\right)$ is by definition independent of the number of capabilities

present in a country $\left(k_{c,0}^{a}\right)$. Moreover, we notice that because of the structure of the Leontief operator, the probability that a country with a given number of capabilities will produce a product requiring x capabilities $\left(\pi\left(c\left(k_{c,0}^{a}\right) \rightarrow p\left(k_{p,0}^{a}=x\right)\right)\right)$ does not depend on the structure of the matrices $\boldsymbol{C}_{ca}$ and $\boldsymbol{P}_{pa}$. After these two considerations, we can differentiate eqn. (16) with respect to $k_{c,0}^{a}$ to show that the diversification of a country increases convexly with the number of capabilities it has. The first two derivatives between the diversification of a country and the number of capabilities present in it are:

$$\frac{dk_{c,0}^{p}}{dk_{c,0}^{a}}=\sum_{x=0}^{N_a} x\left(\frac{k_{c,0}^{a}}{N_a}\right)^{x-1} N_p(k_{p,0}^{a}=x) \tag{38}$$

$$\frac{d^2k_{c,o}^{p}}{dk_{c,0}^{a\ 2}}=\sum_{x=0}^{N_a} x(x-1)\left(\frac{k_{c,0}^{a}}{N_a}\right)^{x-2} N_p\left(k_{p,0}^{a}=x\right). \tag{39}$$

Eqn (39) shows that the second derivative of diversification with respect to the number of capabilities available in a country is always positive, proving that the convex increase in diversification that is associated with the accumulation of complementary capabilities is a result that does not depend on any assumption regarding which countries (products) have (require) what capabilities.

Figure 5 a allows us to understand the implications of this convexity better by plotting eqn. (20) for N_a=50 and a range of q values. It is clear from Figure 5 a that eqn. (20) becomes more curved as *q approaches 1*. This curvature has important implications, as the slope of the $\overline{k_{c,0}^{p}}(k_{c,0}^{a})$ curve represents the number of new products that will become accessible for a country to make or export after increasing the number of capabilities it has by a small amount. As q approaches 1, i.e., as the fraction of capabilities required by the average product increases, the number of new products that become accessible after accumulating a few capabilities becomes small to negligible for countries with only a few capabilities, while at the same time, it becomes extremely large for countries with many of them. This implies that in worlds in which products are intricate, meaning that they require a large fraction of the total number of capabilities that exist ($q \rightarrow 1$), catching up becomes more difficult, since the intricacy of products causes, simultaneously, negligible returns for countries with few capabilities and large returns for countries with many capabilities.

To clarify this point, consider a world in which products require on average 30 capabilities out of 50. Countries with only five capabilities get no returns for the accumulation of one or two extra capabilities. Moreover, they could have no returns for the accumulation of 20 or even 25 capabilities, since there is no guarantee that the capabilities that they accumulate are exactly those required by less intricate products. In the same world, however, countries with 40 or more capabilities will have large returns for the accumulation of any additional one, since it will be possible to put any new capability that comes along into use in combination with the capabilities already present in that country. Hence, the model predicts the existence of a poverty trap, or more accurately a quiescence trap (a trap of development stasis), that will become more pronounced as products become more intricate. In other words, the model therefore predicts that the level of diversification of countries will tend to diverge as products become more intricate.

The model also predicts another road to the quiescence trap which involves increasingly more specific capabilities. Figure 5 b shows equation 20 for $q=0.2$ and N_a values in the range [1 1000]. This figure shows that as the number of capabilities in the world increases, the returns to the accumulation of new capabilities become increasingly more convex showing that the quiescence trap explained above can also emerge as a consequence of the existence of a large number of specific capabilities.

In the next section, we calibrate the model to the data and explore whether the empirically observed data is more consistent with a world in which the quiescence trap is either weak or strong.

Calibration

It is interesting to explore the possible set of parameters that can account for the structure of the country-product network $\boldsymbol{M}_{cp}$. In this section, we calibrate the model using the empirical data by using two different criteria. First, we consider the density of the network (η), which is the ratio between the number of links in the network (i.e. the number of 1's in the matrix) and the total number of possible links ($N_c \times N_p$). For the three datasets considered, we find that the network densities range between 29% and 8% (see Table 1).

	NAICS (N_c=172,N_p=326)	SITC4 (N_c=129,N_p=772)	HS6 (N_c=232,N_p=5109)
η (R*=1)	13.56%	13.53%	8.54%
η (R*=0.5)	28.82%	19.62%	12.57%

Table 1 Network density for the three datasets at R^*=1 and R^*=0.5.

Since all countries are *ex ante* identical in the binomial model, the density (η) of the $\boldsymbol{M}_{cp}$ matrix can be calculated simply as the average fraction of products made by a country, which is that country's diversification divided by the total number of products in the world. Since on average, products require a number of capabilities equal to qN_a, the fraction of products made by a country, and the density of the network (η) are identical and given by:

$$\eta = r^{qN_a}. \qquad (40)$$

Equation (40) defines a constraint between r, q, and N_a ensuring that the networks in the model have the same number of links as the networks in the data. Equation (40) also reduces the number of free parameters in the model from three to two. Since we can always normalize all measures of diversification and ubiquity ($k^p_{c,n}$ and $k^c_{p,n}$) by the number of products and countries in the system (N_p, N_c). Hence, the model has only two free parameters, which we choose as the probability that a country has a capability r and the total number of capabilities in the system N_a. Given that the density (η) of the empirical matrix is known, for a given pair of values for N_a and r, q will be given by:

$$q = \frac{1}{N_a}\frac{\ln(\eta)}{\ln(r)}. \qquad (41)$$

First, we calibrate r and N_a by looking at the empirical relationship between the diversification of a country and the average ubiquity of its products (the $k_{c,0}$-$k_{c,1}$ diagram). We do this by using equation (41) to substitute for q in equation (29), and fit the resulting expression to the empirical

data. Figures 6 and 7 present these fits and show where in the phase space defined by r and N_a the best fits are found. Fits were determined using least squares, for our three datasets and two cut-offs (R^*=0.5 and 1.0).

What is interesting about Figures 6 and 7 is that there is not a single pair of r and N_a values that fits the data best, but rather a well-defined region in the N_a-r phase space where a continuum of similarly good fits are found. Moreover, the set of N_a and r values where the best fits are found is similar for all datasets, suggesting that the model tends to resemble the world only for relatively high values of r and a wide range of N_a and that this part of the calibration does not seem to depend on the level of aggregation of the data.

While geometrically, the $k_{c,0}$-$k_{c,1}$ diagram can be fit equally well considering either a large or a small number of capabilities, the economic interpretation of fits with few capabilities and fits with many capabilities is quite different. To illustrate this, we consider two extreme cases for a matrix of a given density, which for illustrative purposes we take to be η=15%. First, we consider a world with a relatively small number of capabilities (N_a=10), and a probability of a country having a capability of r=0.7 (which is close to where the best fits are found for a small number of capabilities). In this case, a density η=15% is obtained for q=0.53, which means that products on average require 53% of all capabilities in the world, needing on average, five to six capabilities to be made. We call this extreme the ***Concentrated World***, because this is a world in which the number of products being produced is constrained by the relatively large number of capabilities that products require, compared to the number of capabilities available in a country (q/r ~ 75%, the average country has 30% more capabilities than the ones required by the average product).

On the other extreme, we consider the other regime where good fits of the $k_{c,0}$-$k_{c,1}$ diagram are found, in which the probability that a country has a capability is r=0.85. In this case q=0.023, and products require on average less than 3% of the total number of capabilities available in the system. This amount to a total capability requirement of between 11 and 12 capabilities per product (notice that, in absolute numbers, these are more capabilities than in the Concentrated World). We call this second extreme the ***Diluted World***, because in this regime products require a relatively small proportion of the capabilities that countries have (q/r ~ 2.7%, the average country has 3600% more capabilities than those required by the average product). In the Diluted World, the number of products that countries can make is constrained, not by the number of capabilities that products require, but by their specificity. In our example of the Diluted World, countries have, on average, 425 out of 500 capabilities, whereas products require only 11 or 12 capabilities, thus the ability of a country to produce a product is limited by the specificity of these capabilities. In other words, in the Concentrated World countries use almost all of their capabilities in every product, whereas in a Diluted World countries produce products using only a small subset of the total number of capabilities at their disposal. In both worlds the total number of country product connections is the same. Moreover, both the Concentrated and the Diluted Worlds are characterized by relatively deep quiescence traps, albeit for different reasons. In the Concentrated World, the quiescence trap is deep because products require a large fraction of capabilities, whereas in the Diluted World the quiescence trap is deep because the capabilities that products require are very specific.

Fitting the model to the $k_{c,0}$-$k_{c,1}$ diagram defines a band of possible values of r and N_a that are compatible with this particular stylized fact, but do not inform us whether the data resembles either the Concentrated or Diluted World. In order to sort out this dilemma, we consider a second criterion for calibration which is based on the implications of both of these worlds on an important empirical observable associated with the $\boldsymbol{M}_{cp}$ matrix, which is the probability that two products will be co-exported. In a concentrated (diluted) world, many (few) products will tend to be co-exported because they share (do not share) a similar set of requisite capabilities.

Following Hausmann and Klinger (2006) and Hidalgo et al. (2007), we define the proximity between a pair of products p and p' ($\phi_{pp'}$) as the minimum of the pairwise conditional probabilities that a country exports product p if it exports product p'. Since the probability that two products require some of the same capabilities is relatively large in the Concentrated World, we expect the distribution of the conditional probabilities of pairs of products to be co-exported to be relatively high in the Concentrated World. On the other hand, since the probability that two products require some of the same capabilities is small in the Diluted World, we expect co-exports to be relatively rare in the Diluted World, biasing proximities towards smaller values. Hence, we can gauge whether the empirical data resemble the Concentrated or Diluted Worlds by comparing the empirical distribution of proximities with the distribution of proximities that emerges from the model. We do this by numerically implementing the model for a range of r and N_a values and by computing the proximity between products, for the empirical data and the model. Finally we use a weighted Kolmogorov-Smirnov Goodness of Fit Test to find the values of N_a and r, in which the model produces theoretical proximity distributions that more closely resemble the empirical distributions of proximities.

Figures 8 and 9 summarize the result of this exercise by showing – for each one of the datasets and two thresholds – the summary of the Kolmogorov-Smirnov test and a comparison between the empirical and theoretical proximity distributions functions, both in terms of frequency and cumulative distributions. First, we observe that for all datasets and threshold there is a region of the parameter space in which the model accurately reproduces the empirically observed proximity distributions. More important, however, is that the band of values for which the model can reproduce the empirical distribution partially overlaps the band where the best fits for the $k_{c,0}$-$k_{c,1}$ diagram are found. The intersection between both of these regions defines a narrow range of parameters for which the model most accurately describes the empirically data (Figure 10), suggesting that the world lies somewhere in the transition between the *Concentrated* and *Diluted* World, and that both calibration criteria are met using 23 to 25 capabilities, in the most aggregate dataset (NAICS), and 65 to 80 capabilities in the most disaggregate ones (Figure 10 and Table 2).

DATASET	N_a	r	q (from (42))	Kolmogorov-Smirnov Ubiquity distribution	Kolmogorov-Smirnov Diversification distribution
NAICS R*=0.5	25	0.88	0.3893	0.1023	0.2225
NAICS R*=1	23	0.76	0.3165	0.1404	0.1642
SITC-4 R*=0.5	65	0.86	0.1661	0.0849	0.3962
SITC-4 R*=1	80	0.87	0.1795	0.1189	0.3204
HS-6 R*=0.5	70	0.89	0.2542	0.1035	0.4025
HS-6 R*=1	70	0.89	0.3016	0.1084	0.3368

Table 2 Summary of Calibration

We consider as significant the fact that the model adjusts the data only for double-digit values for the number of capabilities because much of economic theory has assumed that the number of factors of production is small. But the structure of world exports suggests that it is very difficult to make sense of these empirically observed patterns unless we assume a much larger number of inputs into the production function.

Next, we compare the empirical data with predictions that the binomial model makes regarding the distributions of diversity and ubiquity (equations (36) and (37)). We take the model into the data by comparing the distributions that emerge from the model for the values of r and N_a determined from the calibration procedure explained above and summarized in Table 2.

Figure 11 compares the empirical ubiquity distribution (probability that a product will be exported by x countries) with the prediction from the binomial model (eqn. (37)), showing that the binomial model accurately reproduces the ubiquity distributions observed in the world. It is worth highlighting that no parameters were readjusted when fitting these distributions. We use the N_a and r values determined from the calibration presented above, making the accuracy of the fits remarkable.

Figure 12 compares the empirical diversification distribution (probability that a country will export x products) with the predictions of the binomial model. Here, however, the binomial model does not represent an accurate fit of the distribution. Indeed, Figure 12 shows that in the empirical data we find a relatively large number of countries with diversifications that are much smaller than those predicted by the binomial model, and a few countries with diversifications that are considerably larger. Since in the binomial model all countries have the same probability of having a given capability, we take this result as evidence of a large degree of inequality in the capability endowment of countries.

Indeed, we can test whether the mismatch between the empirical and analytical distributions of diversification presented in Figure 12 can be corrected by introducing more variation into the number of capabilities that countries have. We can relax the assumption that the probability that a country has a capability is the same for all countries, while leaving the general framework presented above untouched. We do this by using eqn. (33) to estimate the number of capabilities that a country is expected to have given its diversification and the parameters found in the previous calibration. We interpret this value as r_c, the probability that country c has a capability, and use this value to reconstruct numerically $\boldsymbol{C}_{ca}$ and implement the model. Figure 13 compares the distributions of diversification found in the data and the average distribution that comes out of the model after 1,000 numerical simulations. Figure 13 shows that, after incorporating the heterogeneity of the capability endowment of countries into the model, it more accurately reproduces the distribution of diversification observed in the data, supporting the intuition that the observed cross-country heterogeneity in production patterns is a consequence of differences in the capability endowment of countries.

Moreover, this result shows that differences in the capability endowment between countries are more heterogeneous than what would be expected from the binomial model. This heterogeneity also suggests that the quiescence trap is in place, since we can interpret the large fraction of poorly diversified countries as those that have remained quiescent due to the

coordination problems involved in the accumulation of capabilities. We can also interpret the disproportionately large fraction of highly diversified countries as those with a diverse and self-perpetuating capability endowment, due again to the same coordination problem, which in the case of diversified countries becomes an asset instead of a liability. Hence, the empirically observed heterogeneous cross-section can be understood as an expected consequence of the dynamics of the model presented above.

Finally, we use the values obtained from our calibration to estimate the functional form of the relationship between the fraction of capabilities that a country has, and the number of products that it can make. Figure 14 shows our findings demonstrating that, for all of our datasets, the functional form that emerges is one in which the quiescence trap is relatively deep. Interestingly, we find that the curves estimated using the more disaggregated datasets are more curved, indicating a deeper trap. This could suggest that accounting for the production of more disaggregated varieties requires more specific inputs and that our heretofore coarse view of what goes into a product might be tied to the coarseness of our traditional measures of output.

All in all, the results show that economic divergence can emerge simply from input complementarities and that, when these complementarities are taken into account, economic divergence becomes expected, rather than surprising.

Concluding remarks

In this paper, we have studied the characteristics of the relationship between products and the countries that make them. We have found that countries differ not just in how diversified they are, but also in the ubiquity of the products they export. We document a relationship between these two concepts, which is systematic and cannot be explained by the distribution of diversification of countries and the distribution of the ubiquity of products, but speaks to a more fundamental link between the two.

We provide a model for such a link, assuming that each product requires a varied and potentially large set of different complementary non-tradable inputs, which we call capabilities. Countries differ in the number of the capabilities that are present in their territory while products differ in the number of capabilities they require. As a consequence, countries with more capabilities will be more diversified, and products that require more capabilities will be accessible to fewer countries, and hence will be less ubiquitous. This makes the average ubiquity of products exported by countries with more capabilities lower than the average ubiquity of products exported by countries with few capabilities. Empirically, our approach can be used to measure the complexity of a country's capability endowment and the intricacy of products by extending these concepts to include, not only the average ubiquity of a country's exports, but also a series of averages that helps correct for the imperfections of diversification and ubiquity as indicators of the numbers of capabilities in a country or required by a product (Hidalgo and Hausmann 2009). Here, however, we concentrate on formalizing a theory that can help account for some simple and fundamental structural properties of $\boldsymbol{M}_{cp}$, and show that there are deep theoretical implications in the structure of these matrices that emerge from assuming only that

countries are endowed with capabilities and that products require specific combinations of capabilities.

The theory presented above predicts traps in the process of economic diversification. Countries with few capabilities will be able to make few products and will have scant benefits from accumulating any individual additional capability. This is because the likelihood that a new capability will be able to synergize with existing capabilities and become useful for the production of a new product is low in the absence of the other requisite capabilities. Therefore, the demand for any randomly selected new capability is likely to be small in countries with few capabilities. By contrast, countries with many capabilities would be able to produce many new products by combining any new capability with different subsets of the capabilities they already possess. In other words, the model generates increasing returns in terms of diversification to the accumulation of capabilities. Moreover, these increasing returns become more acute as products become more intricate (i.e., as products require a larger proportion of possible capabilities) or when the total number of capabilities in the world is large.

Thinking of products as combinations of capabilities and of countries as incomplete sets of capabilities implies that there are increasing returns to the accumulation of capabilities and the convexity of this relationship increases with the complexity of products, predicting increased divergence between complex and simple economies as the world becomes more intricate. This convexity is different in kind from that described by Rosenstein-Rodan (Rosenstein-Rodan 1943) and formalized by Murphy, Schleifer, and Vishny (Murphy, Shleifer and Vishny 1989). In those models, the question is one of complementarity among few industries that exhibit economies of scale, such as manufacturing and railroads, creating a coordination problem on that margin. A big aggregate push would focus the economy on the good equilibrium by coordinating the supply with the demand for trains at a high level. In this world, the central planning problem appears relatively simple.

In contrast, in the world described in this paper, there are dozens of capabilities, an exponentially growing set of possible combinations between them (2^{Na} to be exact), and incompleteness of the capability set. In such a world, the likelihood that a "big push" will succeed, understood as the provision of a given capability, is lower precisely in the places where the development challenge (or the quiescence trap) is largest, since there will likely be many other missing capabilities that go into making any particular product. In another sense, the model helps clarify the ideas in Hirschman (Hirschman 1958) regarding the creation of disequilibria that would promote backward and forward linkages. In our language, a forward linkage involves the provision of a capability that would then promote the development of an additional product. A backward linkage would be the effective demand for a new capability that emerges from the attempt to make a new product that needs it. Here, forward and backward linkages are the paths towards increasing the variety of capabilities and products. However, the quiescence trap means that this dynamic is more challenging the lower the number of initial capabilities.

When we calibrate the model to our three different datasets, we find that it is hard to make sense of the observed features of the $\boldsymbol{M}_{cp}$ matrix unless we assume that the number of potential capabilities is in the double digits: between 23 and 80. This is quite far from the low

number of inputs that are usually assumed in most production functions, and in particular in previous discussions of the "big push."

If our interpretation of the features of the $\boldsymbol{M}_{cp}$ matrix is correct, countries differ not just in the quantity of each capability, but in the variety of capabilities they have. Seen from this perspective, the challenge of development involves solving the coordination problem between the accumulation of additional capabilities and the demand for those capabilities, which presumes the presence of all the complementary capabilities that would be required by a new activity.

The results presented in (Hidalgo, Klinger, Barabasi and Hausmann 2007) and in (Hausmann and Klinger 2006) argue that comparative advantage evolves by countries moving from existing goods to "nearby" or related goods in The Product Space, suggest that proximity is related to the similarity of the requisite vector of inputs and that production evolves by minimizing the coordination problem. However, the ability to add a product to the production set of a country depends not only on how close a given product is to an already existing one, but also on how many other capabilities are present in the country and used in other, potentially more distant, products.

The description of the development process that emanates from this paper suggests an important distinction between two processes. Diversification may increase because entrepreneurs find valuable new combinations of already existing capabilities. This can be interpreted as the process that is highlighted by Hausmann and Rodrik (Hausmann and Rodrik 2003), where development is seen as a process of discovering the domestic cost of new products. If a new product happens to have a low cost it is because the requisite capabilities were already present. In this paper, we have assumed away this process since we suppose that all feasible products are made and emphasized the coordination problem that is involved in the simultaneous provision of a new capability with the potential demand that emanates from the new products that would become feasible once this capability is present.

This paper opens up several areas of further work. First is the question of why the $\boldsymbol{M}_{cp}$ matrix is triangular rather than diagonal. We have skirted the problem by assuming away the question of quantities, prices, and profits and by having countries produce all the goods for which they have the requisite inputs. Explaining why countries tend to diversify rather than specialize as they develop is an interesting question. For example, the Kremer model (Kremer 1993), which assumes skill complementarities, would generate a diagonal, not a triangular $\boldsymbol{M}_{cp}$ matrix.

Second, is the question of which specifically are the capabilities that seem to go into production. Are they personal skills, organizational outcomes such as non-tradable products and government services, more social outcomes such as rules and norms, or complex combinations of other products such as the capacity to sell over the Internet? Much of the recent discussion has emphasized highly composite measures of institutions which are themselves aggregations of many other more fundamental capabilities, such as the capacity to control corruption, enforce contracts, maintain political stability, regulate economic activity, etc.

Third, it would be interesting to explore how the complexity of an economy evolves. What does it take to accumulate new capabilities? How in practice does this occur? Are non-tradable sectors instrumental in creating the domestic demand for capabilities that can then be redeployed in the production of tradables? Did import substitution policies facilitate or hinder the accumulation of capabilities? Does the new tendency to globalize production facilitate development and the accumulation of capabilities by reducing the set of capabilities that need to be present in a given place for production to occur? Is the bi-modal nature of the world's income and diversification distributions a consequence of the convex relationship between capabilities and products and the existence of a quiescence trap?

Fourth, if more complex products require more diversified inputs, then more complex industries should locate only in urban settings where many inputs are available, thus affecting the distribution of urban diversification and of the complexity of the industries they can support. Do we observe a similar $\boldsymbol{M}_{cp}$ matrix when looking across localities within a country?

In sum, going from assuming a few factors of production to a potentially large and incomplete set of capabilities has implications for some features of the world that are empirically observable and presents some of the old questions of economic development in a somewhat different hue.

FIGURES

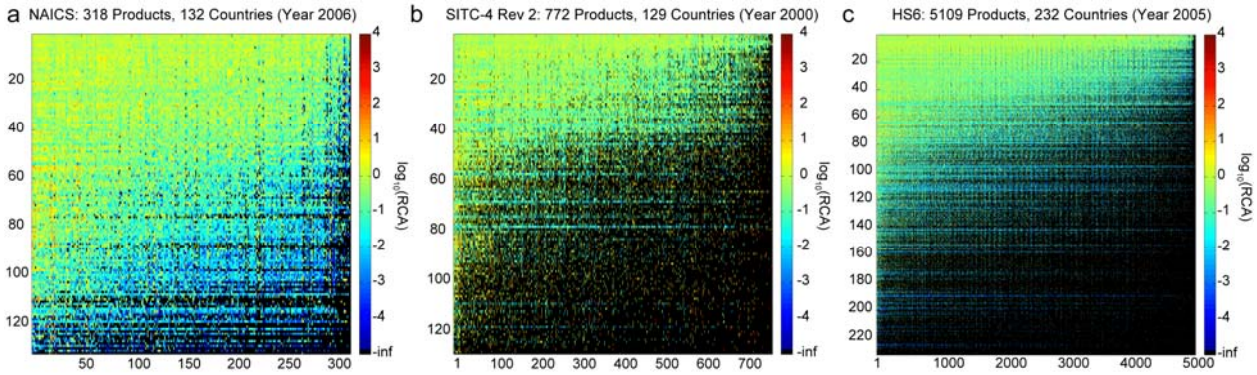


Figure 1 RCA Matrices sorted by the diversification of countries and the ubiquity of their exports calculated using R^*=0.5. The color in each entry of these matrices represents the logarithm, in base 10, of the RCA that countries (rows) have on products (columns). The mapping between colors and values can be read from the colorbar **a** The NAICS dataset for the year 2006 **b** The SITC-4 dataset for the year 2000. **c** The HS-6 dataset for the year 2005.

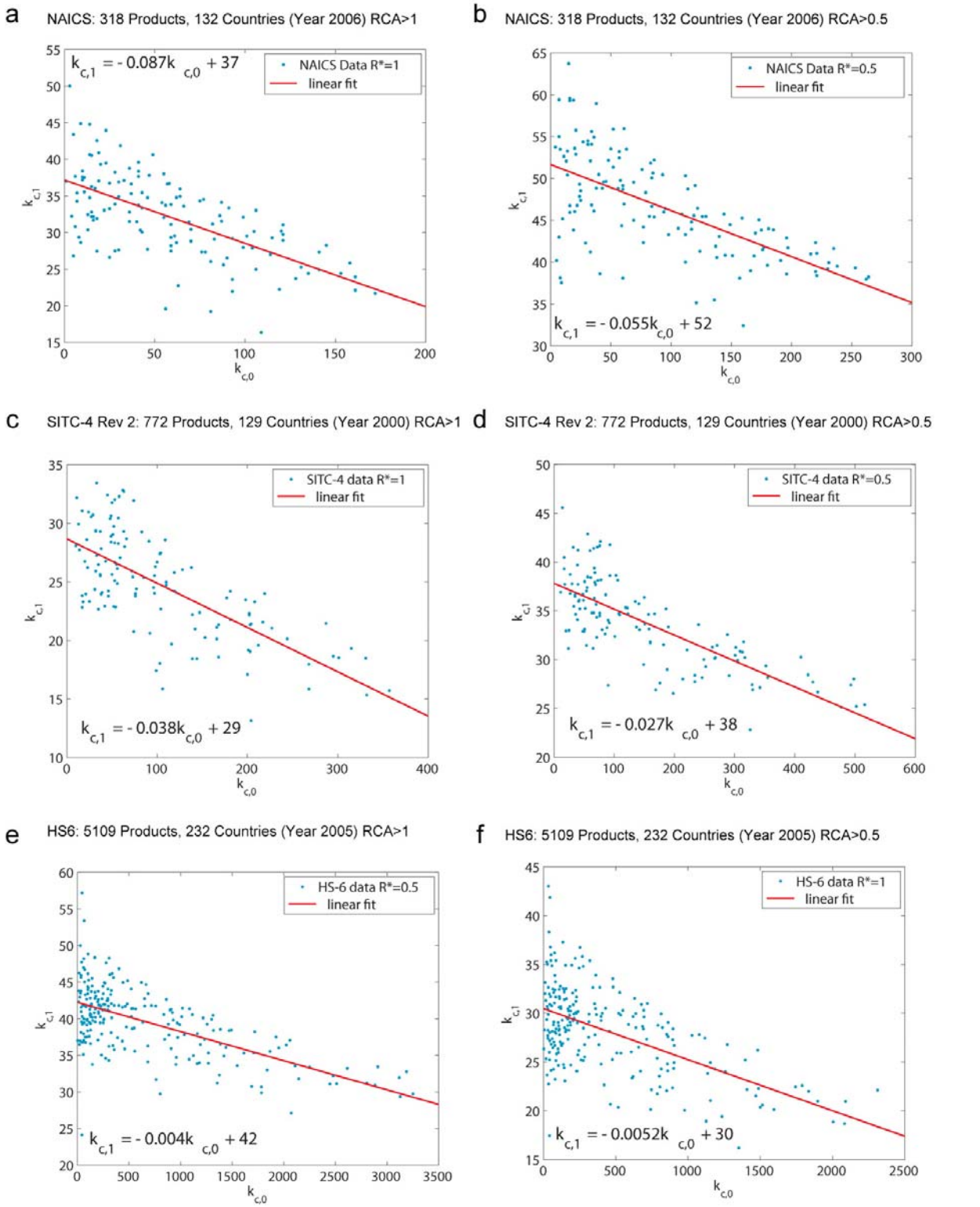


Figure 2 Illustrations of the $k_{c,0}$-$k_{c,1}$ diagrams calculated for R*=1 and R*=0.5 for **a** The NAICS dataset for the year 2006 and R*=1. **b** The NAICS dataset for the year 2006 and R*=0.5. **c** The SITC-4 dataset for the year 2000 and R*=1. **d** The SITC-4 for the year 2000 and R*=0.5. **e** The HS-6 for the year 2005 and R*=1. **f** The HS-6 for the year 2005 and R*=0.5.

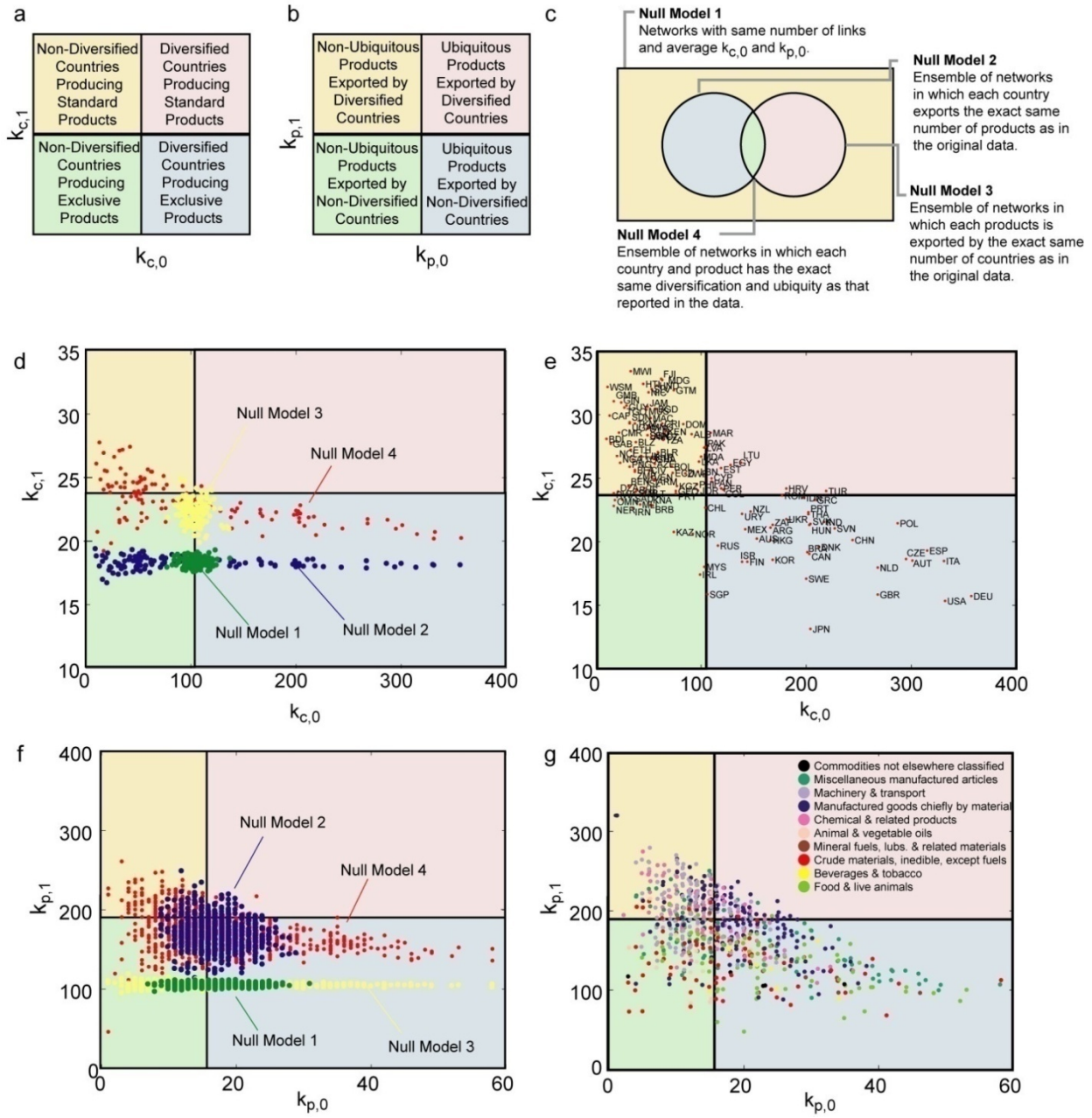
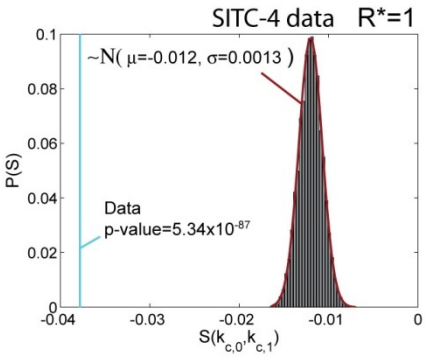


Figure 3 Bipartite network diagrams and their null models for the SITC-4 dataset using R*=1. **a** Schematic explanation of the $k_{c,0}$-$k_{c,1}$ space. **b** Schematic explanation of the $k_{p,0}$-$k_{p,1}$ space. **c** Venn diagram showing the set relations between the four introduced null models. **d** $k_{c,0}$-$k_{c,1}$ diagram for an instance of each of the four null models. **e** Empirically observed $k_{c,0}$-$k_{c,1}$ diagram **f** $k_{p,0}$-$k_{p,1}$ diagram for an instance of each of the four null models **g** Empirically observed $k_{p,0}$-$k_{p,1}$ diagram.



Example of the *p*-value estimation procedure.

The slope of 1000 null models is used to fit a distribution and find the probability of observing any given slope. This is then used to calculate the probability of observing the slope characterizing the diagram observed in the data given each null model.
This example illustrates the procedure for the $k_{c,0}$-$k_{c,1}$ diagram, null model 4 and the SITC-4 dataset with a threshold of R*=1.

Significance of the $k_{c,0}$-$k_{c,1}$ diagram

significant at: ◆ 1 % ▲ 5% ● 10%

		Statistical Significance (p-value smaller than) Null Model			
Dataset	R*	1	2	3	4
HS-6	1	◆ 1E-26	◆ 1E-100	▲ 4E-02	◆ 2E-35
HS-6	0.5	◆ 4E-15	◆ 1E-100	● 1E-01	◆ 5E-14
SITC-4	1	◆ 5E-17	◆ 1E-100	◆ 6E-04	◆ 5E-87
SITC-4	0.5	◆ 9E-22	◆ 1E-100	◆ 1E-02	◆ 3E-47
NAICS	1	◆ 6E-39	◆ 1E-100	◆ 8E-04	◆ 1E-22
NAICS	0.5	◆ 4E-30	◆ 1E-100	▲ 2E-02	◆ 3E-06

Significance of the $k_{p,0}$-$k_{p,1}$ diagram

		Statistical Significance (p-value smaller than) Null Model			
Dataset	R*	1	2	3	4
HS-6	1	◆ 1E-100	◆ 1E-100	◆ 3E-86	◆ 6E-14
HS-6	0.5	◆ 1E-100	◆ 1E-100	◆ 9E-36	◆ 8E-10
SITC-4	1	◆ 1E-100	◆ 1E-100	◆ 2E-18	◆ 3E-86
SITC-4	0.5	◆ 1E-100	◆ 1E-100	◆ 5E-07	◆ 7E-22
NAICS	1	◆ 1E-100	◆ 1E-100	◆ 7E-08	◆ 2E-35
NAICS	0.5	◆ 1E-100	◆ 1E-100	◆ 3E-03	◆ 8E-08

Figure 4 Statistical significance of the $k_{c,0}$-$k_{c,1}$ and the $k_{p,0}$-$k_{p,1}$ diagrams.

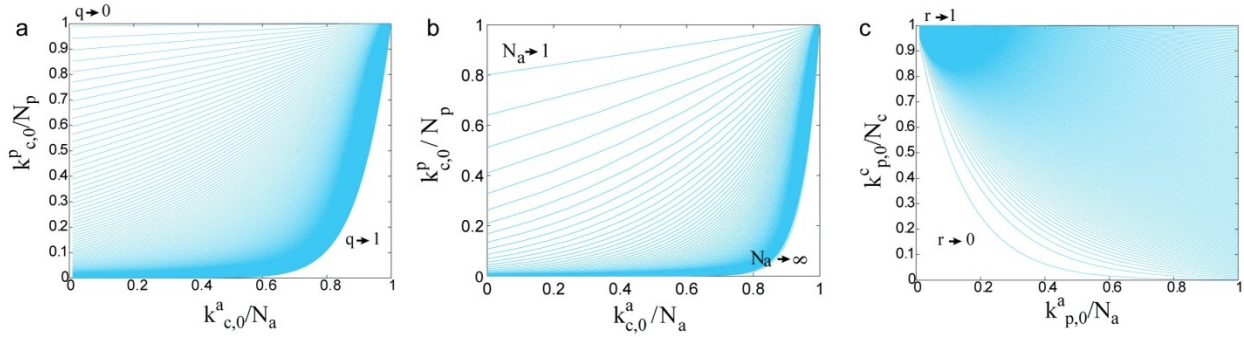


Figure 5 analytically calculated predictions from the Binomial Model. **a** Fraction of all products that countries can make as a function of the fraction of all capabilities that countries have for values of q ranging from 0 to 1 and N_a=50. **b** Fraction of all products that countries can make as a function of the fraction of all capabilities that countries have for q=0.2 and N_a ranging from 1 to 1000. **c** Fraction of countries that make a product as a function of the fraction of capabilities it requires.

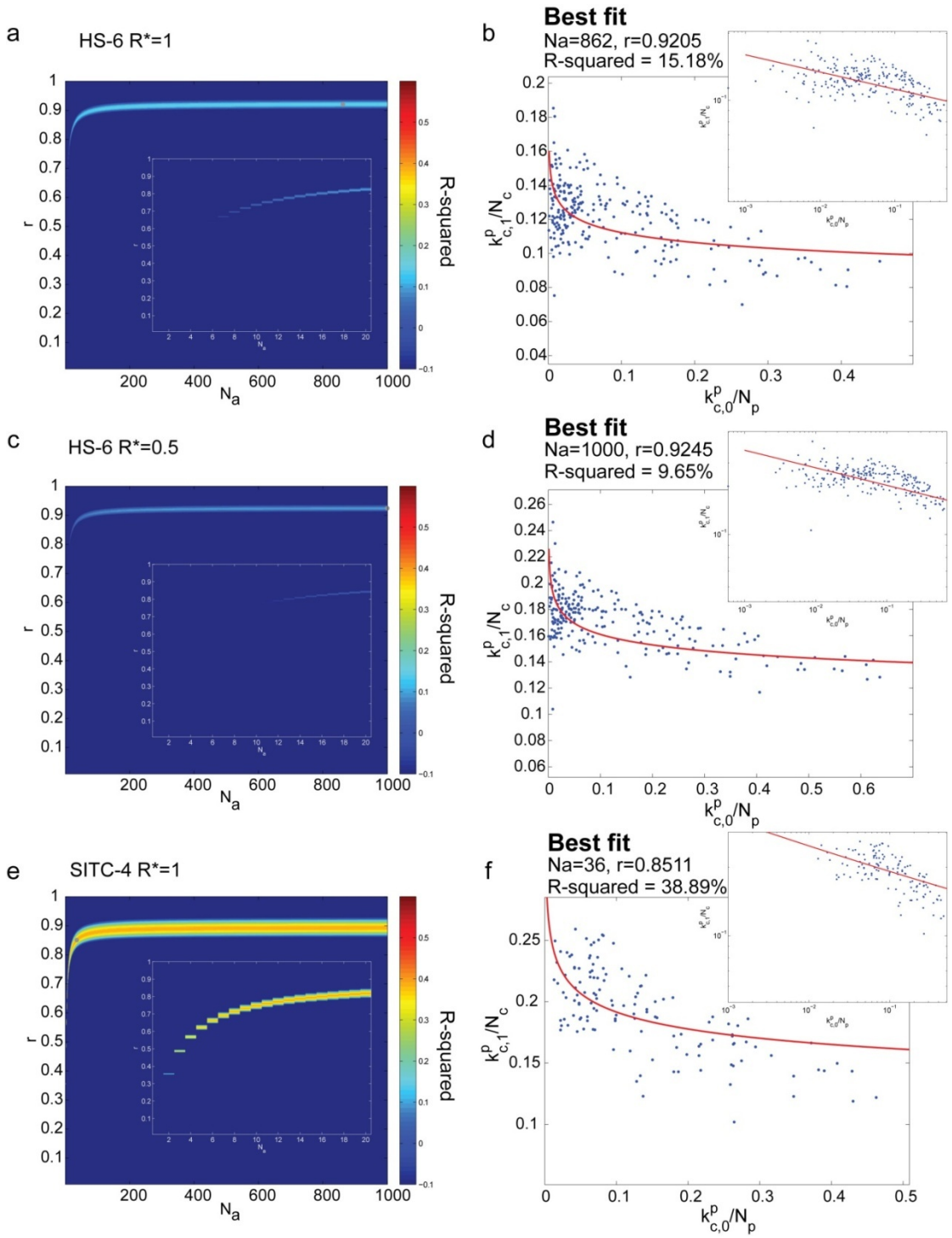


Figure 6 Calibration of the Binomial Model to the $k^p_{c,0}$-$k^p_{c,1}$ diagram. **a** Values of the R-squared statistic of the Binomial Model fit to the $k^p_{c,0}$-$k^p_{c,1}$ for r~[0,1] and N_a~[0,1] for the HS-6 dataset and R*=1. **b** Illustration of the best fit of the binomial model to the HS-6 dataset with R*=1. **c** Same as **a,** but for R*=0.5. **d** Same as **b** but for R*=0.5. **e** Same as **a,** but for the SITC-4 dataset. **f** Same as **b** but for the SITC-4 dataset.

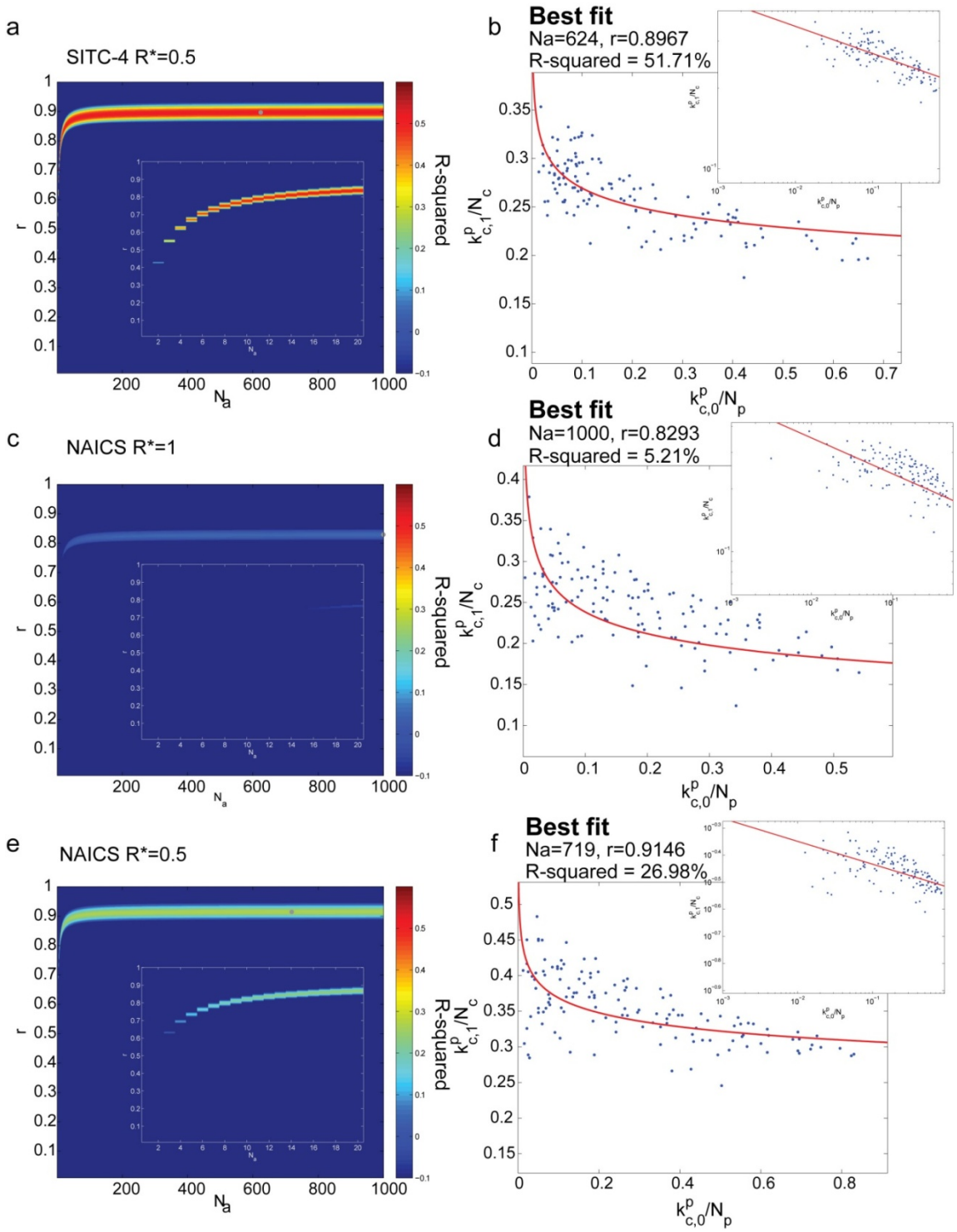


Figure 7 Calibration of the Binomial Model to the $k^p_{c,0}$-$k^p_{c,1}$ diagram. **a** Values of the R-squared statistic of the Binomial Model fit to the $k^p_{c,0}$-$k^p_{c,1}$ for r~[0,1] and N_a~[0,1] for the SITC-4 dataset and R*=0.5. **b** Illustration of the best fit of the binomial model to the SITC-4 dataset with R*=1. **c** Same as **a** but, for the NAICS dataset and R*=1. **d** Same as **b** but for the NAICS dataset and R*=1. **e** Same as **a** but, for the NAICS dataset. **f** Same as **b** but for the NAICS dataset.

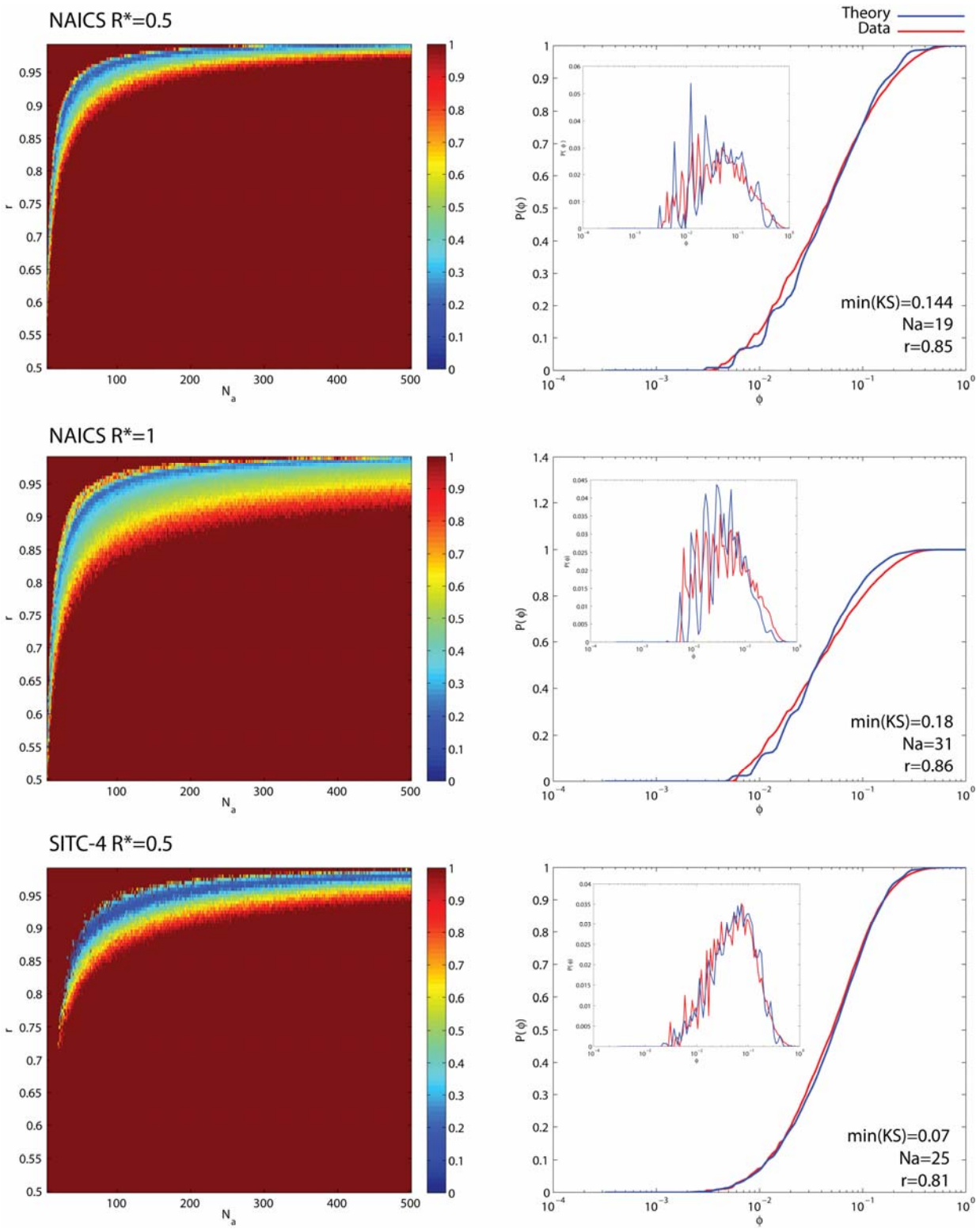


Figure 8 Calibration of the Binomial Model to the Proximity Distribution. The left panel shows the numerically obtained values for the weighted Kolmogorov-Smirnov test used to find the values of r and Na that result in proximity distributions that resemble more closely the empirical data. Low values of the KS test indicate similar distributions. The right panel compares the best fit of the model (blue) with empirical data (red) showing both the cumulative and density probability distributions (inset).

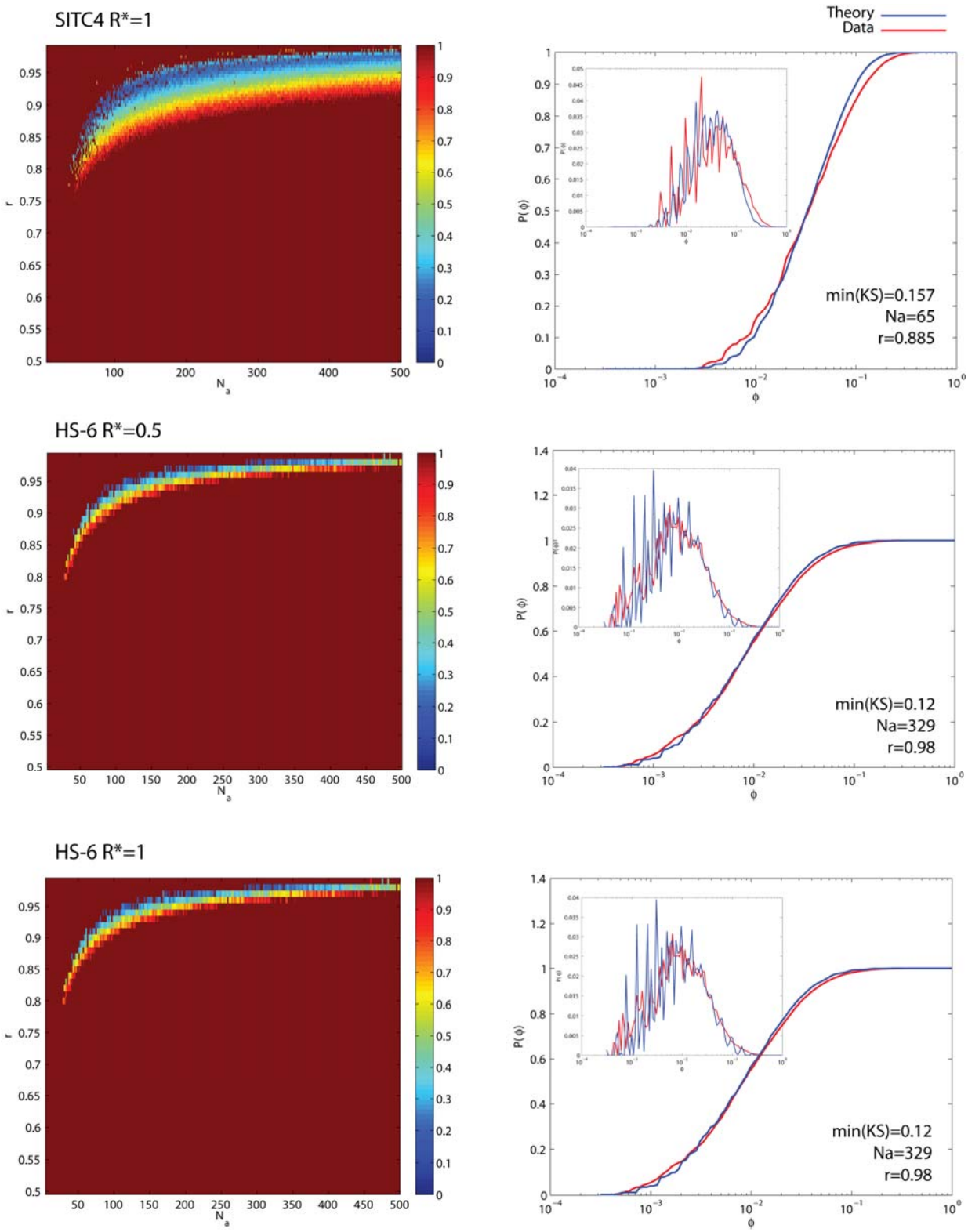


Figure 9 Calibration of the Binomial Model to the Proximity Distribution. The left panel shows the numerically obtained values for the weighted Kolmogorov-Smirnov test used to find the values of r and Na that result in proximity distributions that resemble more closely the empirical data. Low values of the KS test indicate similar

distributions. The right panel compares the best fit of the model (blue) with empirical data (red) showing both the cumulative and density probability distributions (inset).

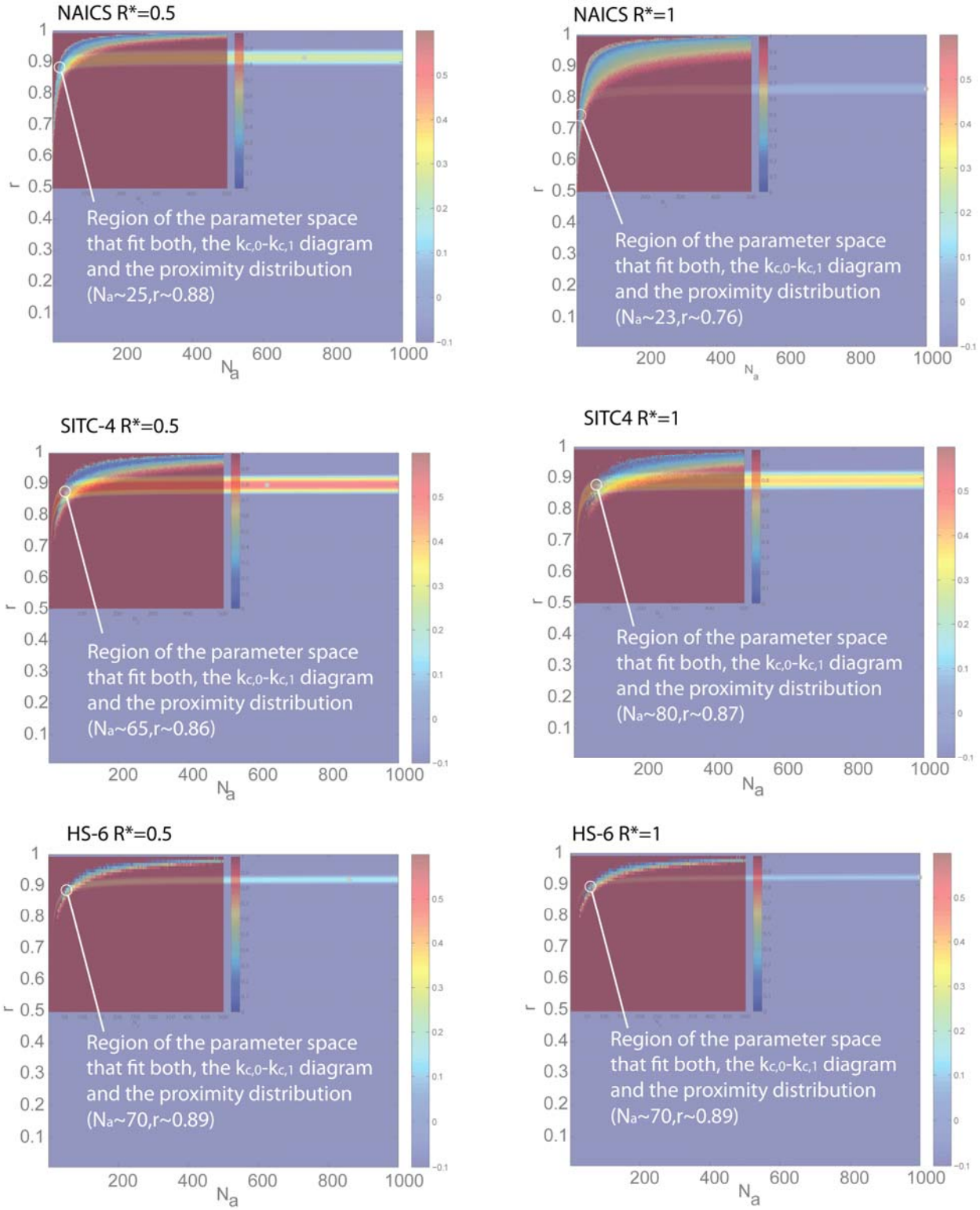


Figure 10 Overlap between both calibration procedures showing the regions of the Na-r parameter space where the model approximates both, the kc,0-kc,1 diagram and the proximity distributions that are observed in the data.

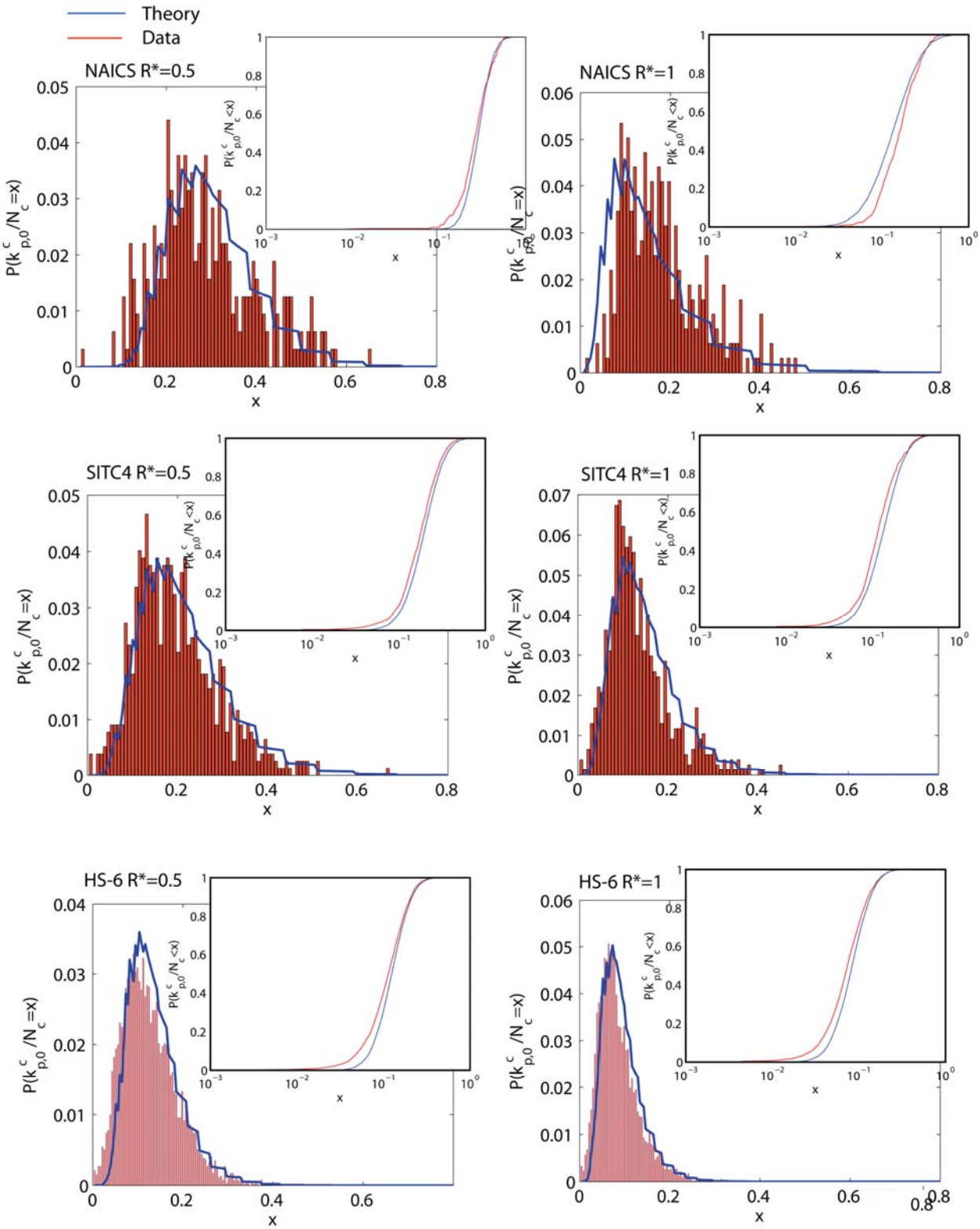


Figure 11 Comparison between the empirically observed ubiquity distribution and the one exhibited by the model for the parameters found by calibrating it to the $k_{c,0}$-$k_{c,1}$ diagram and the proximity distribution. Full figures show density probability functions whereas the insets show their respective cumulative distribution. KS statistics are presented in Table 2.

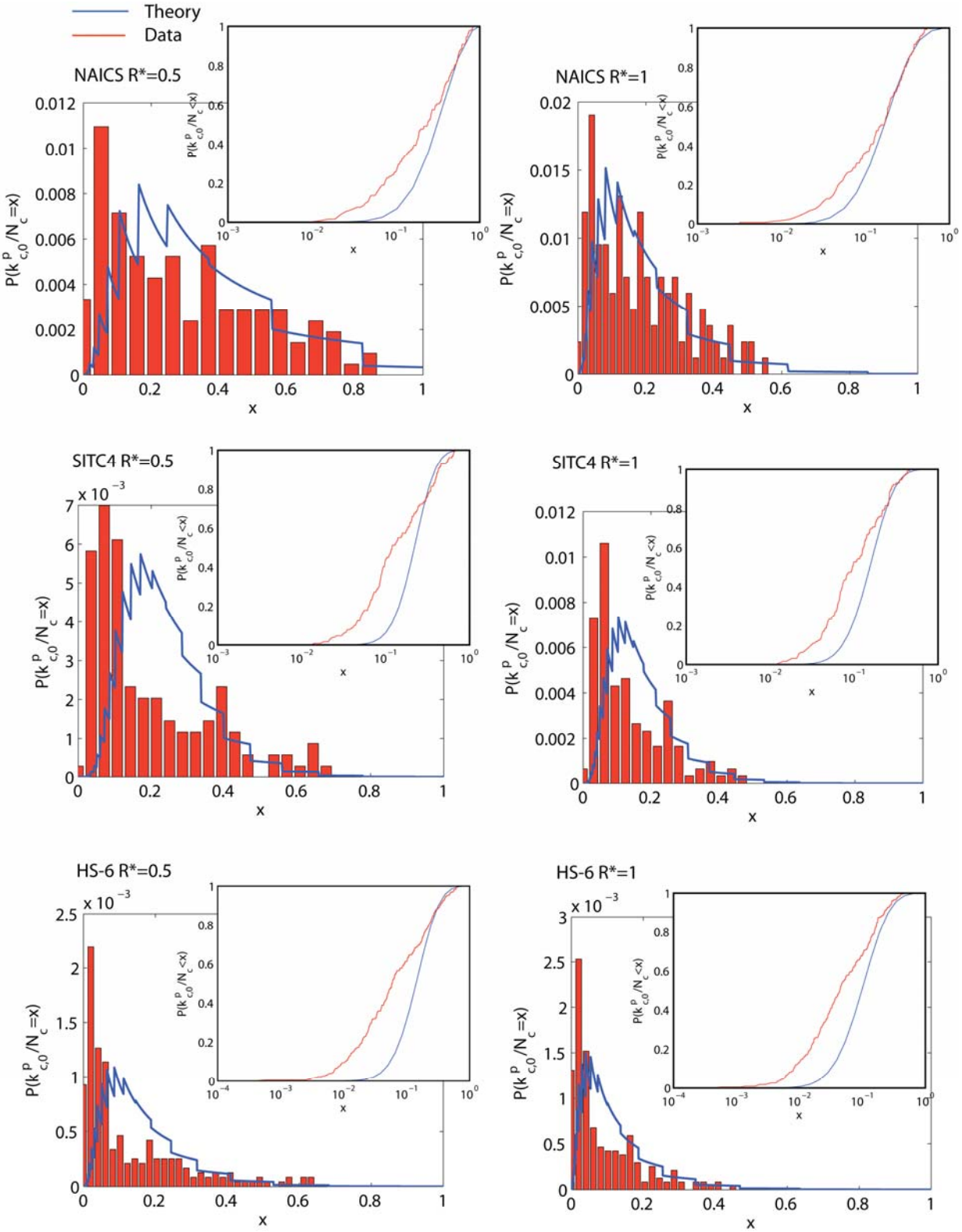


Figure 12 Comparison between the empirically observed ubiquity distribution and the one exhibited by the model for the parameters found by calibrating it to the $k_{c,0}$-$k_{c,1}$ diagram and the proximity distribution. Full figures show density probability functions, whereas the insets show their respective cumulative distribution. Kolmogorov-Smirnov statistics are presented in Table 2.

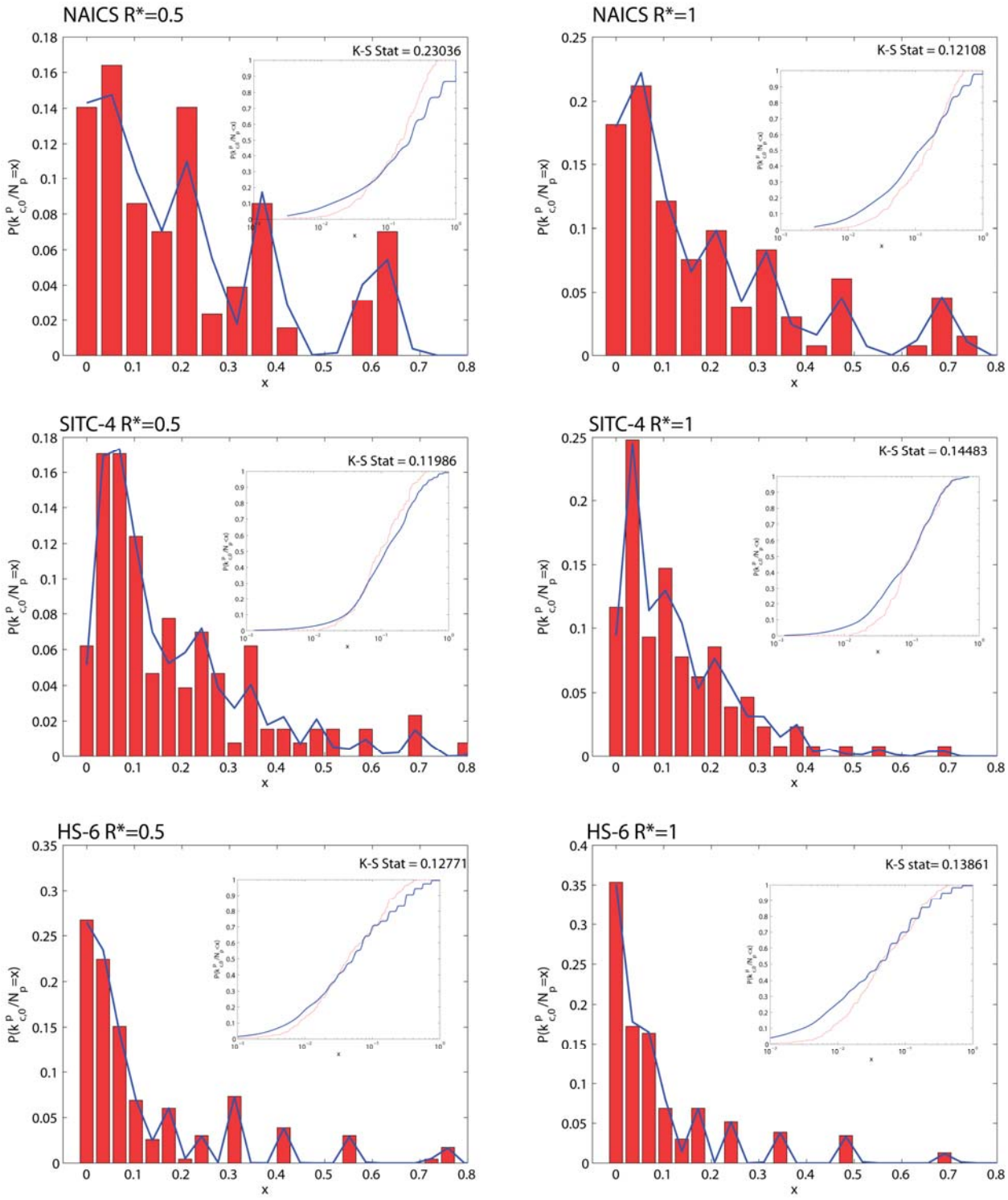


Figure 13 Comparison between the empirically observed ubiquity distribution and the one exhibited by the model for the parameters found by calibrating it to the $k_{c,0}$-$k_{c,1}$ diagram and the proximity distribution. Full figures show density probability functions, whereas the insets show their respective cumulative distribution. Kolmogorov-Smirnov statistics can be read from the figure.

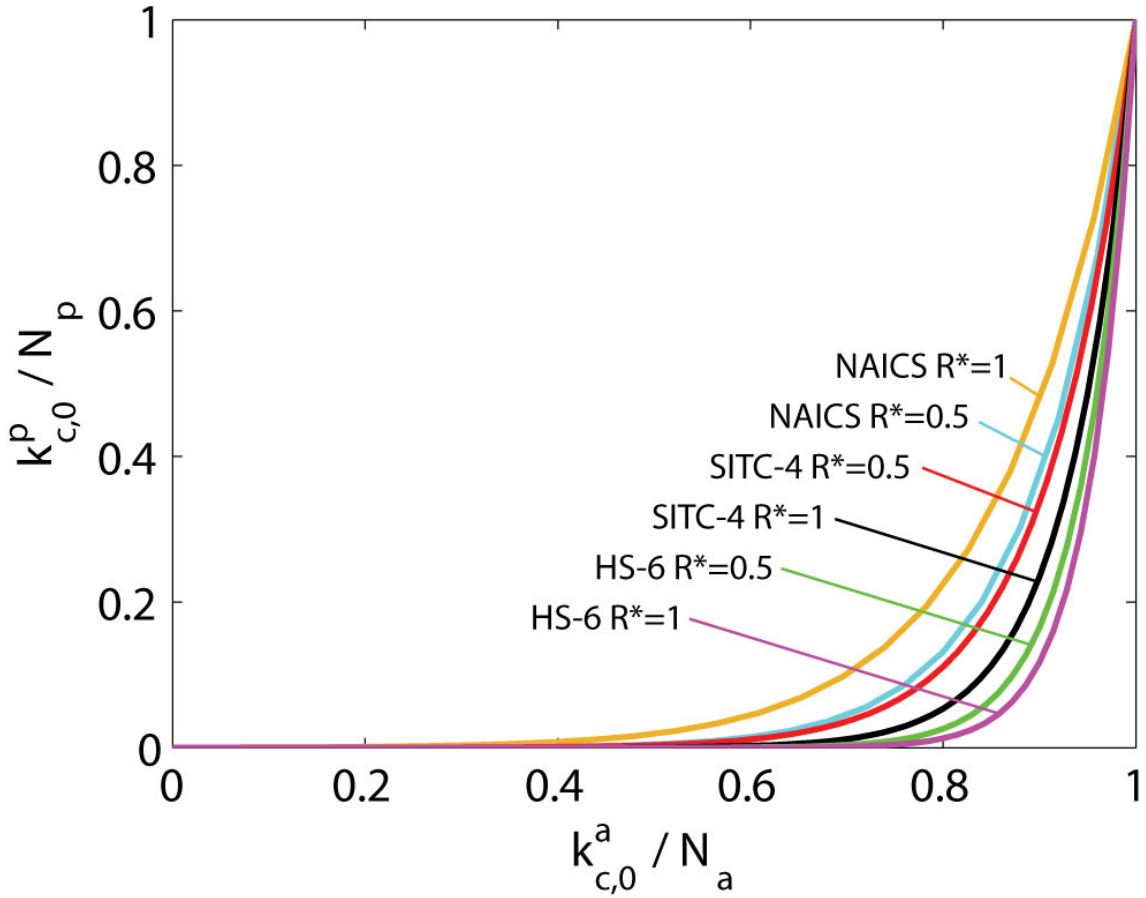


Figure 14 Relationship between fraction of capabilities that countries have (x-axis) and the fraction of products that countries make (y-axis) for the calibrated model values.

REFERENCES

Acemoglu, D., P. Antras, and E. Helpman, "Contracts and technology adoption," American Economic Review, 97 (2007), 916-943.
Balassa, B., "THE PURCHASING-POWER PARITY DOCTRINE - A REAPPRAISAL," Journal of Political Economy, 72 (1964), 584-596.
Dixit, A. K., and J. E. Stiglitz, "MONOPOLISTIC COMPETITION AND OPTIMUM PRODUCT DIVERSITY," American Economic Review, 67 (1977), 297-308.
Feenstra, Robert C., Robert E. Lipsey, Haiyan. Deng, Alyson C. Ma, and Henry. Mo, "World Trade Flows:1962-2000," NBER Working Paper No. W11040, (2005).
Gaulier, G., and S. Zignano, "Title," 2009.
Grossman, G. A., and E. Rossi-Hansberg, "Trading Tasks: A Simple Theory of Offshoring," American Economic Review, 98 (2008), 1978-1997.
Hausmann, R., and B. Klinger, "Title," 2006.
Hausmann, R., and D. Rodrik, "Economic development as self-discovery," Journal of Development Economics, 72 (2003), 603-633.
Helpman, E., and P.R. Krugmann, *Market Structure and Foreign Trade* (Cambridge, MA: MIT Press, 1985).
Hidalgo, C. A., and R. Hausmann, "The building blocks of economic complexity," Proceedings of the National Academy of Sciences of the United States of America, 106 (2009), 10570-10575.
Hidalgo, C. A., B. Klinger, A. L. Barabasi, and R. Hausmann, "The product space conditions the development of nations," Science, 317 (2007), 482-487.
Hirschman, A. O., "THE PATERNITY OF AN INDEX," American Economic Review, 54 (1964), 761-762.
Hirschman, Albert O., *The Strategy of Economic Development* (New Haven, CT: Yale University Press, 1958).
Hummels, D., and P. J. Klenow, "The variety and quality of a nation's exports," American Economic Review, 95 (2005), 704-723.
Jost, L., "Entropy and diversity," Oikos, 113 (2006), 363-375.
Kauffman, Stuart A., *The Origins of Order* (New York, New York: Oxford University Press, 1993).
Kremer, M., "THE O-RING THEORY OF ECONOMIC-DEVELOPMENT," Quarterly Journal of Economics, 108 (1993), 551-575.
Krugman, P., "The Increasing Returns Revolution in Trade and Geography," American Economic Review, 99 (2009), 561-571.
Krugman, P. R., "INCREASING RETURNS, MONOPOLISTIC COMPETITION, AND INTERNATIONAL-TRADE," Journal of International Economics, 9 (1979), 469-479.
Melitz, M. J., "The impact of trade on intra-industry reallocations and aggregate industry productivity," Econometrica, 71 (2003), 1695-1725.
Murphy, K. M., A. Shleifer, and R. W. Vishny, "INDUSTRIALIZATION AND THE BIG PUSH," Journal of Political Economy, 97 (1989), 1003-1026.
Rodriguez-Clare, A., "Clusters and comparative advantage: Implications for industrial policy," Journal of Development Economics, 82 (2007), 43-57.
Rosenstein-Rodan, PN., "Problems of Industrialization of Eastern and South-Eastern Europe," The Economic Journal, 53 (1943), 202-211.

Saviotti, P. P., and K. Frenken, "Export variety and the economic performance of countries," Journal of Evolutionary Economics, 18 (2008), 201-218.
Schott, P. K., "Across-product versus within-product specialization in international trade," Quarterly Journal of Economics, 119 (2004), 647-678.
Smith, A., *An Inquiry into the Nature and Causes of the Wealth of Nations* (Chicago, IL: University Of Chicago Press, (1977) [1776]).
Weitzman, M. L., "Recombinant growth," Quarterly Journal of Economics, 113 (1998), 331-360.