

# Stages of Diversification in a Neoclassical World \*

Catia Batista<sup>†</sup> and Jacques Potin<sup>‡</sup>

Forthcoming *Economic Letters*

## Abstract

Recent research has documented a U-shaped industrial concentration curve over an economy's development path. How far can neoclassical trade theory take us in explaining this pattern? We estimate the production side of the Heckscher-Ohlin model using industry data on 44 developed and developing countries for the period 1976-2000. Decomposing the implied changes in industrial concentration over time shows that at least one third of these changes seems to be explained by a Rybczynski effect. This result suggests that capital accumulation led poor countries to diversify their industrial production, while rich countries made their production more concentrated in highly capital-intensive industries.

**JEL Codes:** F11; L16; O40.

**Keywords:** Economic Growth and International Trade; Heckscher-Ohlin; Diversification; Specialization; Industrial Concentration; Structural Change.

---

\*Corresponding Author: Catia Batista. Mailing Address: Faculdade de Economia, Universidade Nova de Lisboa, Campus de Campolide, Lisboa 1099-032, Portugal. Phone number: +351213801604. Email address: catia.batista@novasbe.pt. We thank an anonymous reviewer, Nicolas Coeurdacier, Laurence Lescourret, Peter Neary, Radu Vranceanu, Adrian Wood, and other participants at a number of conferences and seminars for their comments. Excellent research assistance was provided by Christoph Lakner. We gratefully acknowledge financial support from the George Webb Medley Fund at the University of Oxford and from Nova Forum at Nova University of Lisbon.

<sup>†</sup>Nova School of Business and Economics, Faculdade de Economia - Nova University of Lisbon.

<sup>‡</sup>ESSEC, Paris.

# 1 Introduction

Countries grow through stages of initial industrial diversification and eventual concentration. Indeed, a U-shaped industrial concentration curve over an economy's development path was documented by Imbs and Wacziarg's 2003 American Economic Review article. According to their finding, in a first stage of development, as a country's per capita income increases, the country's manufacturing sector becomes more diversified. In a second stage, while per capita income keeps rising, the country's industry becomes more concentrated.

This phenomenon has major macroeconomic consequences. For instance, Koren and Tenreyro (2007) showed that industrial concentration is important in explaining the volatility of the growth of GDP per capita. This implies that, especially in their first stage of development, regardless of per capita income growth, poor countries with a more concentrated productive structure are potentially left increasingly vulnerable to higher GDP volatility. In this context, understanding the mechanisms that generate the two stages of industrial concentration put forward by Imbs and Wacziarg (2003) becomes all the more relevant for the design of economic policies aimed at reducing income fluctuations, especially in less developed economies.

Perhaps surprisingly, there is to the best of our knowledge no empirical study examining the sources underlying the industrial specialization patterns described by Imbs and Wacziarg (2003).<sup>1</sup> Our paper proposes a simple testable hypothesis: we argue that the U-shaped industrial concentration curve might be a direct result of a Rybczynski effect, whereby countries accumulating capital should initially diversify their industrial production (as they move away from labor-intensive industries to capital-intensive industries), and eventually concentrate in the production of highly capital-intensive goods.<sup>2</sup>

Our work uses the Heckscher-Ohlin (HO) model of international trade as a framework to empirically study the evolution of industrial concentration over time. Using data on 44 developed and developing countries over the 1976-2000 period, we aim at quantifying the importance of factor accumulation in explaining structural change within the industrial sector. For this purpose, the model is first estimated using panel data in a way that is novel but consistent with the mainstream trade empirical literature.<sup>3</sup> The estimation results show that the traditional

---

<sup>1</sup>In their analysis, these authors described four potential theoretical mechanisms that, when combined, could explain the observed empirical patterns. Namely, they justified industrial diversification as the likely result of portfolio diversification arguments as in Acemoglu and Zilibotti (1997) and of a "taste for variety" implied by non-homothetic preferences *à la* Dixit and Stiglitz (1977). The second stage of industrial concentration could be explained by falling transportation costs, allowing countries to specialize according to their comparative advantages, as predicted by Ricardian models with a continuum of goods like that of Dornbusch et al. (1977), or by increasing returns to scale generated by the clustering of economic activity in specific regions or countries *à la* Krugman (1991).

<sup>2</sup>Acemoglu and Guerrieri (2008) developed a model related to this hypothesis as they also explain the non-balanced nature of growth with a factor-proportions model. However, they model a closed economy and empirical work is not tackled, but only called for in their concluding remarks.

<sup>3</sup>Our estimation strategy closely follows Schott (2003) in allowing for different cones of specialization and aggregating ISIC industries into two "HO aggregates" - one being labor-intensive and the other capital-intensive.

$2 \times 2$  HO model with fixed cone cutoffs explains well specialization patterns at the HO aggregate level over the whole period under analysis.

We then use the estimation results to decompose the evolution in industrial concentration as measured by a Herfindahl index into components that can be attributed to Rybczynski effects. We obtain that capital accumulation is key in explaining the evolution of specialization across HO aggregates - the Rybczynski effect. Our calculations indicate that on average at least one third of the changes in industrial concentration can be explained by this effect - and nearly half of the changes observed in the period 1976-1990.

The paper is organized as follows. Section 2 estimates the production side of the traditional Heckscher-Ohlin model. Section 3 decomposes the changes in industrial concentration over time. The last section summarizes and concludes.

## 2 The production side of the Heckscher-Ohlin model

### 2.1 Theoretical framework

Consider  $N$  countries,  $n = 1, \dots, N$ . At date  $t$ , country  $n$  is endowed with a quantity  $K_n^t$  of capital and a quantity  $L_n^t$  of labor. There are two goods ( $i = l, k$ ) that can be produced in each country using capital and labor. Good  $l$  is labor-intensive, and good  $k$  is capital-intensive. There is no factor intensity reversal. Both goods are produced with constant returns to scale (CRS) by competitive firms, and the marginal product of each factor is positive and decreasing at the firm (and industry) level. All countries have access to the same technology. Each country is small and can freely trade goods on the world market at date- $t$  prices.

In a competitive equilibrium, countries with a low capital-labor ratio ( $K_n^t/L_n^t \in (\tau_0 = 0, \tau_1^t)$ ) specialize in the production of the labor-intensive good. Countries with an intermediate capital-labor ratio ( $K_n^t/L_n^t \in (\tau_1^t, \tau_2^t)$ ) produce the two goods. Countries with a high capital-labor ratio ( $K_n^t/L_n^t \in (\tau_2^t, \tau_3 = +\infty)$ ) specialize in the production of the capital-intensive good.

<Figure 1 about here.>

Figure 1 represents the production patterns implied by the theoretical model.<sup>4</sup> The dashed line represents value-added per worker for the labor-intensive good over a country's "development path". The solid line represents value-added per worker for the capital-intensive good. The *Rybczynski effect* says that, with fixed prices and technologies, capital accumulation in

---

However our estimation is performed using panel data, whereas that of Schott (2003) is based on a cross-section of countries for the year 1990. In addition, we introduce multiplicative country fixed effects that account for country heterogeneity, accounting for cross-country differences in total factor productivity (TFP) as in Treffer (1995), and also for land and natural resource availability. We also make adjustments for factor quality as in Treffer (1993).

<sup>4</sup>See Deardorff (1974, 1999) for further details.

the cone of diversification leads to a reduced production of the labor-intensive good and to an increased production of the capital-intensive good. That is exactly what can be observed in Figure 1: initially only (an increasing quantity of) the labor-intensive good is produced, but once the economy reaches the capital-labor ratio necessary to enter the cone of diversification, this production falls until it reaches zero in the final cone of specialization, where only (an increasing quantity of) the capital-intensive good is produced. The capital-intensive good displays the opposite behavior over this economy's development path, as it continues growing from zero as soon as the economy reaches the cone of diversification. The thin straight line that is tangent to the two curves determines the capital-labor ratios in the two industries (equal to  $\tau_1^t$  and  $\tau_2^t$ ).

<Figure 2 about here.>

The patterns of specialization given by Figure 1 can be used to construct a measure of productive concentration for this economy. Figure 2 depicts the Herfindahl index implied by the HO model.<sup>5</sup> The model's prediction of a U-shaped curve for industrial concentration when a country moves along its "development path", i.e. when it accumulates capital, is evident in this figure - a prediction very much in line with the industrial concentration patterns documented by Imbs and Wacziarg (2003).

## 2.2 Econometric model and estimation strategy

### Linearity assumption

In order to estimate the theoretical model presented in the previous section, we begin by assuming that value-added in the production of each good  $a$  (where  $a = l, k$ ) by country  $n$  at time  $t$  (denoted  $VA_{na}^t$ ), divided by the total number of workers in country  $n$  at time  $t$  (denoted  $L_n^t$ ), is a piece-wise linear function of country  $n$ 's capital-labor intensity at time  $t$  (denoted  $K_n^t/L_n^t$ ), and can be expressed as:

$$\frac{VA_{na}^t}{L_n^t} = \sum_{c=1}^3 \left[ \alpha_{ca}^t + \beta_{ca}^t \frac{K_n^t}{L_n^t} \right] I_{\left[ \tau_{c-1}^t < \frac{K_n^t}{L_n^t} < \tau_c^t \right]} \quad a = l, k \quad (1)$$

where  $\tau_0^t \equiv 0$  and  $\tau_3^t \equiv +\infty$ , and  $c = 1, 2, 3$  denotes the different cones of production, which are delimited by cone cutoffs  $\tau_c^t$  as in the theoretical HO model described in the previous subsection. Note that cones and cone cutoffs are common to all countries, although our econometric model allows for them to potentially vary over time. The parameters describing production in

---

<sup>5</sup>With two goods only, denoting by  $s_l$  the share of the labor-intensive good in total value-added, and by  $s_k$  the share of the capital-intensive good, the Herfindahl index is simply  $HI = s_l^2 + s_k^2$ . It attains a maximum of 1 when only one good is produced, and a minimum of 0.5 when the two shares are equal to 50%.

each cone ( $\alpha_{ca}^t$  and  $\beta_{ca}^t$ ) also vary over time and are subject to several restrictions implied by the theoretical model.<sup>6</sup>

## HO aggregates

Given that in reality many more than two goods are produced, we use the methodology proposed by Schott (2003) to aggregate value-added produced in the various industries in our dataset into two HO aggregates, one produced using labor more intensively and the other produced using capital more intensively. This aggregation procedure is theoretically appropriate because industries are gathered according to their capital-intensity usage as in the production side of the HO model, unlike when using classifications such as the International Industrial Classification (ISIC) which provides collections of goods with similar end use that may be produced in totally distinct ways. Using Schott's (2003) aggregation methodology, we allow for the same ISIC industry to be classified as belonging to a labor-intensive aggregate in a country where capital usage is relatively low (consider the footwear industry in Panama) and as belonging to a capital-intensive aggregate in a country where this industry uses capital in production intensively (consider the Italian footwear industry).<sup>7</sup>

We form HO aggregates as follows. For each country, all the industries with a capital-labor ratio below the HO cutoff  $q$  are grouped into a labor-intensive HO aggregate, and all the industries with a capital-labor ratio above the HO cutoff are grouped into a capital-intensive HO aggregate. In country  $n$  at date  $t$ , the production of the labor-intensive aggregate is therefore  $VA_{nl}^t = \sum_{i: \frac{K_{ni}^{*t}}{L_{ni}^{*t}} \leq q} VA_{ni}^t$ , while the production of the capital-intensive aggregate is equal to  $VA_{nk}^t = \sum_{i: \frac{K_{ni}^{*t}}{L_{ni}^{*t}} > q} VA_{ni}^t$ , where  $K_{ni}^{*t}/L_{ni}^{*t}$  is the effective capital to effective labor ratio defined in the next section. Note that we assume that the HO cutoff does not change over time: this is why  $q$  does not depend on  $t$ .<sup>8</sup>

## International differences in factor quality

Following Treffer (1993), we wish to take into account factor-augmenting international productivity differences. Let  $z_{fn}^t$  be the productivity of factor  $f = K, L$ , in country  $n$  at date  $t$ . If  $K_n^t$  and  $L_n^t$  are, respectively, the measured quantities of capital and labor, then the effective quantities of capital and labor are given by  $K_n^{*t} = z_{Kn}^t \cdot K_n^t$  and  $L_n^{*t} = z_{Ln}^t \cdot L_n^t$ .

<sup>6</sup>Namely, in cone 1,  $\alpha_{1l}^t = 0$  and  $\beta_{1l}^t > 0$ ,  $\alpha_{1k}^t = 0$  and  $\beta_{1k}^t = 0$ . In cone 2,  $\alpha_{2l}^t > 0$  and  $\beta_{2l}^t < 0$ ,  $\alpha_{2k}^t = 0$  and  $\beta_{2k}^t > 0$ . And, finally, in cone 3,  $\alpha_{3l}^t = 0$  and  $\beta_{3l}^t = 0$ ,  $\alpha_{3k}^t > 0$  and  $\beta_{3k}^t > 0$ . Across cones, restrictions imply that  $\tau_1^t \cdot \beta_{1l}^t = \alpha_{2l}^t$ ;  $\alpha_{2l}^t + \tau_2^t \cdot \beta_{2l}^t = 0$  and  $\tau_2^t \cdot \beta_{2k}^t = \alpha_{3k}^t$ .

<sup>7</sup>For any ISIC industry, cross-country differences in factor intensity might reflect the fact that countries specialize in goods that differ in quality. It might also be the result of an international fragmentation of the production process, with the labor-rich countries specializing in the production stages that are labor-demanding. The aggregation procedure we adopt using value-added data circumvents both of these concerns.

<sup>8</sup>In a setting with more than two HO aggregates and cones, each HO aggregate would first see its production increasing and then falling. This would imply that there would be additional dynamics of industrial concentration underlying the ones we uncover at a more aggregate level. In this sense, we may regard our work as an aggregate summary of the industrial dynamics underlying at a more refined sector definition level.

### Country and time fixed effects

Even after accounting for cross-country factor-augmenting productivity differences, there are still cross-country total factor productivity (TFP) differences remaining. We take this possibility into account by incorporating time-invariant Hicks-neutral productivity differences in our empirical model. Estimating this TFP effect incorporates the impact of land or natural resources on productivity.<sup>9</sup> The effect of such productivity differences is a pure scale effect: the production level of each aggregate is multiplied by the same TFP factor. Note that the cone cutoffs do not change, while all the intercepts and slopes are multiplied by this factor.

In order to measure the effect of changes in global technological progress and of aggregate prices, we also introduce multiplicative time dummies  $\delta_a^t$ , specific to each aggregate but common to all countries.

The econometric model we finally estimate is given by:

$$\frac{VA_{na}^t}{L_n^{*t}} = (1 + \delta_a^t)(1 + \delta_{na}) \sum_{c=1}^3 \left[ \alpha_{ca} + \beta_{ca} \frac{K_n^{*t}}{L_n^{*t}} \right] I_{[\tau_{c-1} < \frac{K_n^{*t}}{L_n^{*t}} < \tau_c]} + \epsilon_{na}^t, \quad a = k, l, \quad c = 1, 2, 3. \quad (2)$$

Note that we estimate one national productivity factor for each aggregate ( $\delta_{na}$  where  $a = k, l$ ). We check *ex post* that  $\delta_{nk} = \delta_{nl}$  are of the same order of magnitude for each country.

Further note that we assume that the cone cutoffs do not change over time. This requires that the relative price of the two aggregates does not change or that this change is compensated by biased technological change. The validity of this assumption is also checked *ex post*.

### Estimation procedure

In order to estimate our econometric model, we need to select values for the HO cutoff  $q$  and the two cone cutoffs,  $\tau_1, \tau_2$ . The selection of these parameters was done as follows.

We estimated model (2) using constrained Non Linear Least Squares (constrained NLLS) on our panel dataset over a grid of HO aggregate cutoffs and cone cutoffs. We took values for  $q = 2, 500; \dots; 15, 000$ , with steps of 2,500, and took values for  $\tau_1 = 500; \dots; 10, 000$ , with steps of 500, and  $\tau_2 = 3, 000; \dots; 40, 000$ , also with steps of 500).

The total Sum of Squared Residuals (SSR) was then computed for each set  $(q, \tau_1, \tau_2)$  by summing the SSRs for both HO aggregates. We selected the triplet  $(q, \tau_1, \tau_2)$  such that the total SSR was minimized.

---

<sup>9</sup>See Schott (2003) for a short discussion and Leamer (1987) for a more advanced treatment. We do not explicitly include land and natural resources in our empirical analysis for two reasons. First, these factors are likely time-invariant or nearly so, and therefore included and indistinguishable from other country specific, time invariant factors included in the country fixed effects in our estimation. Second, because we examine only industrial production and not all sectors (notably the primary sector), land and natural resources are not likely to play a major role in explaining the dynamics of industrial concentration.

## 2.3 Data and construction of the main variables

This paper uses data from the United Nations Industrial Development Organization (UNIDO, 2005 a,b) and the Penn World Table version 6.1 (PWT) (Heston et al., 2002). The UNIDO data set (revision 2 at the 3-digit level) presents data for 28 sectors, but several countries aggregate data for two or more sectors (like "food products" and "beverages") into a larger one. To appropriately recognize missing data and to make data comparable across countries, we follow Koren and Tenreyro (2008) and aggregate sectors in order to obtain a consistent classification across countries. This leaves us with the 19 sectors or industries. To estimate the model, we use 5-year averages for the periods 1976-1980, 1981-1985, 1986-1990, 1991-1995, and 1996-2000. We use the data for a country in a given period when they are available and when the country is open to international trade (according to Sachs and Warner, 1995).<sup>1011</sup>

### Value-added

Value-added is expressed in 1996 U.S. dollars, computed using the exchange rates and national deflators from the PWT. We do not use PPP-adjusted values: we assume that all goods produced with the same capital-labor ratio can be sold at the same price on international markets, as in the HO model.<sup>12</sup>

### Capital

To compute the stocks of capital at the industry level, we use investment data from the UNIDO at the 3-digit (UNIDO, 2005a) and 4-digit (UNIDO, 2005b) levels. The first database provides data for the period 1963-2001, and the second database covers the period 1985-2001. The second database contains data missing in the first one. The two matrices were merged.<sup>13</sup> We took investment in current U.S. dollars, and used exchange rates from the PWT to translate these numbers in the current national currency. Implicit national investment deflators were computed as the ratio value of national investment in current national currency units (ICUR) / value of national investment in 1996 national constant prices (IKON). We then used the perpetual inventory method to compute stocks of capital at the sectoral level.<sup>14</sup>

---

<sup>10</sup>Imbs and Wacziarg (2003) results hold regardless of country openness. The assumption of country openness is not necessary and is only made to bring the data closer to the assumptions of the neoclassical trade model.

<sup>11</sup>A list of all sectors, geographic coverage of key variables, and industry capital-labor ratios are available from the authors upon request.

<sup>12</sup>Value-added did not need to be expressed in dollars for the purpose of our estimation or decomposition exercise. This conversion is only done so that cross-country comparisons of value-added are possible.

<sup>13</sup>Even after data merging, the database contains many lacunae. In order to compute stocks of capital, we had to make assumptions about these missing values. When there are gaps within a sequence and when there are fewer than six consecutive years of missing data, the sequence was completed using a linear interpolation. When the beginning (end) of a sequence is missing, we replaced the last (first) three missing values with an average of the first (last) three available values.

<sup>14</sup>The exact formula is the one used by Leamer (1984, p. 233). We kept capital stock estimates only when we have at least eight consecutive years of investment data (once gaps are filled, as explained in the previous

In order to adjust capital stocks for international productivity differences, we rely on Eaton and Kortum (2001) estimates for the price of equipment goods for a given country, which can be interpreted as the real price of an *effective* unit of equipment good. As prices are normalized so that the U.S. price is equal to one, we can find for each country a proxy for the productivity of equipment goods (relative to the U.S.) by computing the inverse of the reported price of capital. As a result, we use the measured price of equipment capital to adjust reported investment *downward* when investment is *costly*.

## Labor

To measure the quantity of labor used in a production in a given industry, we use data from the UNIDO database. These data are corrected for educational differences using the method proposed by Hall and Jones (1999)<sup>15</sup> and the data on educational achievement from Barro and Lee (2000).

## 2.4 Estimation results

The  $(q, \tau_1, \tau_2)$  triplet with best fit is given by  $(q = 7, 500; \tau_1 = 3, 000; \tau_2 = 21, 000)$ . Table 1 presents the results of our estimation for this cutoff triplet. It shows that the Rybczynski effect is substantial and highly statistically significant. In the cone of diversification, countries that accumulate capital produce fewer and fewer of the labor-intensive goods ( $\hat{\beta}_{1l} > 0$  implies  $\hat{\beta}_{2l} < 0$ ) and more and more of the capital-intensive goods ( $\hat{\beta}_{2k} > 0$ ). As a result, capital-accumulating countries have produced more and more of the capital-intensive goods.

<Table 1 about here.>

The period dummies appear to be small and not statistically significant. There is little change in the absolute price of each HO aggregate, or any price change has been almost exactly compensated by a biased technological progress. As a result, the relative price of the two HO aggregates does not seem to vary significantly over the 1976-1995 period. We interpret this finding as a validation of our assumption that the cone cutoffs do not move too much over time between 1976 and 2000. This estimation result has important implications for our analysis in that it points to the great dominance of the Rybczynski effect over other potential Heckscher-Ohlin effects such as relative price changes or technological change.

note).

<sup>15</sup>The correction factor for labor in country  $n$  at date  $t$  is  $\pi_{ln}^t = \phi(s_n^t) = s_{n0}^t + s_{n2}^t e^{2 \times 0.134} + s_{n4}^t e^{4 \times 0.134} + s_{n6}^t e^{4 \times 0.134 + 2 \times 0.101} + s_{n8}^t e^{4 \times 0.134 + 4 \times 0.101} + s_{n10}^t e^{4 \times 0.134 + 4 \times 0.101 + 2 \times 0.068} + s_{n12}^t e^{4 \times 0.134 + 4 \times 0.101 + 4 \times 0.068}$ , where  $s_{ne}^t$  is the fraction of the population over 25 with  $e$  years of education.



### 3 The U-shaped industrial concentration curve in a neo-classical world

Our estimation of the Heckscher-Ohlin model points to the importance of the Rybczynski effect in explaining the dynamics of industrial specialization over time for our panel of 44 developed and developing countries over 1976-2000. We take this finding as a first indication that factor accumulation might indeed explain the U-shaped concentration curve described by Imbs and Wacziarg (2003), according to which countries grow through two stages of diversification: a first stage of development in which, as a country's per capita income increases, the country's manufacturing sector becomes more diversified; and a second stage in which, while per capita income keeps rising, the country's industry becomes more concentrated.

The hypothesis we examine is that poor labor-abundant countries that accumulate capital diversify their output by moving away from labor-intensive industries to capital-intensive industries, while rich capital-abundant countries that accumulate capital concentrate in the production of highly capital-intensive goods. As discussed in Section 2.1. and illustrated by Figure 2, this hypothesis is theoretically capable of producing Imbs and Wacziarg's (2003) empirical U-shaped concentration curve. In this section, we evaluate the extent to which this theoretical hypothesis can explain the empirical observation.

We should however note that the decomposition exercise we carry out makes use of manufacturing data only - and while this is closer to the spirit of the neoclassical trade model, we need to recognize that, even for highly skilled and technologically-advanced service sectors, it is not likely that Rybczynski effects are capable of documenting the U-shaped stages documented by Imbs and Wacziarg (2003) outside of manufacturing.

#### 3.1 Decomposing the change in the Herfindahl index

In order to measure the concentration of a country's manufacturing sector, we use a Herfindahl Index (HI) defined for country  $n$  at date  $t$  as  $HI_n^t = \sum_{u=1}^S \left( \frac{VA_{nu}^t}{\sum_{v=1}^S VA_{nv}^t} \right)^2$ , with  $VA_{nu}^t$  denoting the value-added of sector  $u$  and  $S$  the total number of sectors in the economy.

To isolate the effect of the neoclassical factors, we proceed as follows. With two HO aggregates, we can decompose the HI as:

$$HI_n^t = (s_{nl}^t)^2 HI_{nl}^t + (s_{nk}^t)^2 HI_{nk}^t$$

with

$$s_{na}^t = \frac{\sum_{i=1}^{S_a^t} VA_{nai}^t}{\sum_{j=1}^S VA_{nj}^t} \quad \text{and} \quad HI_{na}^t = \sum_{i=1}^{S_a^t} \left( \frac{VA_{nai}^t}{\sum_{j=1}^{S_a^t} VA_{naj}^t} \right)^2 \quad a = l, k$$

The total HI can therefore be expressed as a weighted sum of the HIs at the HO aggregate level, with the weights being equal to the squared shares of the aggregates. It is straightforward to show that the change in concentration can be expressed as:

$$\Delta HI_n^{t,t+1} \equiv HI_n^{t+1} - HI_n^t = \underbrace{\sum_{a=l,k} s_{na}^2 (HI_{na}^{t+1} - HI_{na}^t)}_{\text{within-aggregate change}} + \underbrace{\sum_{a=l,k} [(s_{na}^{t+1})^2 - (s_{na}^t)^2] HI_{na}}_{\text{between-aggregate change}}, \quad (3)$$

with

$$s_{na}^2 = \frac{s_{na}^{t^2} + s_{na}^{t+1^2}}{2} \quad \text{and} \quad HI_{na} = \frac{HI_{na}^t + HI_{na}^{t+1}}{2} \quad a = l, k$$

We now present two hypothetical extreme cases to provide simple intuition about this decomposition. In both cases, there are two aggregates with two industries in each aggregate.

### Case 1: Between effect only

Under this scenario, it is assumed that over time an economy's value-added increases evenly in the two capital-intensive industries, whereas it remains constant in the two labor-intensive economies. In this situation, value-added in the four industries changes as in the following example:

$$(VA_{l_1}^t, VA_{l_2}^t, VA_{k_1}^t, VA_{k_2}^t) = (10, 10, 10, 10) \rightarrow (VA_{l_1}^{t+1}, VA_{l_2}^{t+1}, VA_{k_1}^{t+1}, VA_{k_2}^{t+1}) = (10, 10, 20, 20).$$

Concentration does not change within each aggregate, and therefore there is no within-aggregate change, only a between-aggregate effect. We have  $\Delta HI_n^{t,t+1} = 0.278 - 0.250 = 0.028$ . The concentration index increases as the shares of the two aggregates become more unequal.

This between-effect can be described as a "Heckscher-Ohlin effect": it happens as a result of a shift of production to a set of goods that use a factor intensively. Such a shift can be explained by the neoclassical trade model and its direct extensions: Rybczynski effects, biased technological progress, change in relative prices, or a decline in transportation costs combined with comparative advantages *à la* HO.

### Case 2: Within effect only

Under this alternative scenario, it is assumed that the value-added of one industry in each aggregate is growing at an even rate, while the value-added in the other industry in each aggregate remains constant. Suppose that value-added changes as follows:

$$(VA_{l_1}^t, VA_{l_2}^t, VA_{k_1}^t, VA_{k_2}^t) = (10, 10, 10, 10) \rightarrow (VA_{l_1}^{t+1}, VA_{l_2}^{t+1}, VA_{k_1}^{t+1}, VA_{k_2}^{t+1}) = (10, 20, 10, 20).$$

The share of each aggregate does not change: there is no between-aggregate effect, only a within-aggregate effect. The Herfindahl index changes by  $\Delta HI_n^{t,t+1} = 0.278 - 0.250 = 0.028$ ,

exactly the same magnitude as before. However, the concentration index now rises because each aggregate becomes more concentrated. A within effect can be due to a decline in transportation costs combined with comparative advantages *à la* Ricardo, or with economies of scale as studied in the new economic geography literature.

### Compositional changes in the HO aggregates

In reality, a combination of between-aggregate and within-aggregate effects is likely to occur. In addition, most countries in our sample accumulated capital over time, and national industries moved from the labor-intensive aggregate to the capital-intensive one. As a result, the composition of the aggregates also changed over time. We control for this composition effect by keeping constant the composition of the HO aggregates between the base and final periods.

Note that movements of ISIC industries towards the capital-intensive aggregate over time can also be attributed to Rybczynski effects within the HO framework: in this case, capital accumulation leads not only to a change in the mix of goods that are produced (towards goods that use capital more intensively in production), but also to an alteration in the technology that is used in production, which becomes more capital-intensive – up to a level where the good starts to be classified as capital-intensive – a result that ends up showing in the data as a movement from the production of the labor-intensive aggregate to the production of the capital-intensive aggregate, fitting the classical definition of Rybczynski effect.

In this sense, when we do not exclude compositional effects we are actually capturing two types of Rybczynski effects: one that leads to an increase in the production of capital-intensive goods that were classified as such to begin with (the Rybczynski effect that remains after excluding compositional effects), and a second Rybczynski effect whereby observed increases in the production of capital-intensive goods are also due to the fact that some goods that were previously produced using labor-intensive technology are now being produced using capital-intensive goods. The fact that we are able to quantitatively isolate compositional effects allows us to measure the relative importance of these two types of Rybczynski effects, an original contribution to the literature.<sup>16</sup>

### Results of the decomposition

Using the methodology just proposed, we decompose changes of industrial concentration over 10-year intervals. Namely, we consider changes in the Herfindahl index ( $\Delta HI$ ) and its between ( $BE$ ) and within components ( $WE$ ) over three moments in time: 1976-1980, 1986-1990, and 1996-2000. Tables 2 and 3 show results for both unadjusted and adjusted (for compositional

---

<sup>16</sup>The assumption of only two aggregates in our model was made for derivation simplicity reasons. Estimating the HO model with a larger number of sectors (and cones) would imply that each HO aggregate would first see its production increase and then fall over an economy's development path, as it accumulates capital. This would still imply the existence of Rybczynski effects, but there would be additional dynamics of industrial concentration underlying the ones we uncover at a more aggregate level. In this sense, we may regard our exercise as an aggregate summary of the industrial dynamics underlying at a more refined sector definition level.

effects) data.

<Table 2 about here.>

<Table 3 about here.>

Our results point to neoclassical factors explaining a large fraction of observed changes in industrial concentration, especially between 1976 and 1990. Over that period, the between effect explained an average of 46.1% of the (absolute) magnitudes of changes in industrial concentration, while it explained an average 36.6% of changes between 1986 and 2000.

The between effect is positive for most countries, indicating a positive contribution to industrial concentration. Moreover, countries with negative between effects (denoting diversification of the industrial productive structure) are mostly less developed (such as Bolivia, Hong-Kong, Malaysia, and the Philippines in the earlier period, and Jordan, India, Indonesia, and Panama in the more recent period), which conforms to the theoretical predictions of the Heckscher-Ohlin model.

If we account for changes in the composition of HO aggregates, thereby focusing only on the very strict definition of Rybczynski effect discussed in the previous section, the weight of between effects on total changes in industrial concentration decreases to 29.7% and 26.9%, respectively for the periods 1976-1990 and 1986-2000. We observe an average reduction in the relative contribution of the between effect to the overall HI change due to the compositional adjustment of HO aggregates. This is evidence of significant movements of ISIC industries across HO aggregates over time - in particular, of industries moving to the HO capital-intensive aggregate as their capitalistic intensity increases.

It must be noted that our empirical approach does not take into account changes in specialization within HO aggregates due to changes in factor proportions - as would be the case of a move towards the production of capital-intensive goods within the capital-intensive HO aggregate when there is capital accumulation.<sup>17</sup> For that reason, our results (even when not excluding compositional effects as discussed above) should be interpreted as lower bounds to the total concentration/diversification effect of factor accumulation. Nevertheless, it is clear that there is still room for other types of explanations for changes in patterns of industrial diversification and specialization.

### 3.2 The shares of the HO aggregates and the Rybczynski effect

We turn now to measuring the fraction of between-aggregate changes in industrial specialization that can be attributed to factor accumulation - as opposed to other neoclassical factors (such

---

<sup>17</sup>This is a type of intra-industry specialization discussed in detail by Batista and Potin (2013).

as changes in relative prices or biased technological change).<sup>18</sup> We start by computing the value-added per worker for aggregate  $a$  in country  $n$  predicted by our estimates, which allows calculating predicted shares for both aggregates.

$$\left(\frac{VA_{na}^t}{L_n^{*t}}\right)_{HO} = (1 + \widehat{\delta}_a^t)(1 + \widehat{\delta}_{na}) \sum_{c=1}^3 \left[ \widehat{\alpha}_{ca} + \widehat{\beta}_{ca} \frac{K_n^{*t}}{L_n^{*t}} \right] I_{[\widehat{\tau}_{c-1} < \frac{K_n^{*t}}{L_n^{*t}} < \widehat{\tau}_c]}, a = k, l. \quad (4)$$

The predicted share of aggregate  $a = k, l$  is then computed as:

$$\widehat{(s_{na}^{t+1})}_{HO} = \frac{\left(\frac{VA_{na}^{t+1}}{L_n^{*t+1}}\right)_{HO}}{\left(\frac{VA_{nl}^{t+1}}{L_n^{*t+1}}\right)_{HO} + \left(\frac{VA_{nk}^{t+1}}{L_n^{*t+1}}\right)_{HO}}. \quad (5)$$

We can then compute the change in shares for both aggregates explained by our model:

$$\widehat{(s_{na}^{t+1})}_{HO} - \widehat{(s_{na}^t)}_{HO}.$$

To isolate the role played by factor accumulation (the Rybczynski effect) in determining diversification-concentration patterns, we begin by computing an estimate of the value-added per worker for aggregate  $a$  in country  $n$  that would have been observed without any change in the coefficients of the model or in the country residuals, but with changes in  $K_n/L_n$ :

$$\left(\frac{VA_{na}^t}{L_n^{*t}}\right)_{RYB} = \sum_{c=1}^3 \left[ \widehat{\alpha}_{ca} + \widehat{\beta}_{ca} \frac{K_n^{*t}}{L_n^{*t}} \right] I_{[\widehat{\tau}_{c-1} < \frac{K_n^{*t}}{L_n^{*t}} < \widehat{\tau}_c]}, a = k, l. \quad (6)$$

The new predicted share of aggregate  $a$  is given by:

$$\widehat{(s_{na}^{t+1})}_{RYB} = \frac{\left(\frac{VA_{na}^{t+1}}{L_n^{*t+1}}\right)_{RYB}}{\left(\frac{VA_{nl}^{t+1}}{L_n^{*t+1}}\right)_{RYB} + \left(\frac{VA_{nk}^{t+1}}{L_n^{*t+1}}\right)_{RYB}}. \quad (7)$$

We then compute the change in shares for both aggregates explained by the Rybczynski effect only:

$$\widehat{(s_{na}^{t+1})}_{RYB} - \widehat{(s_{na}^t)}_{RYB}.$$

Figures 3 and 4 report the actual changes in the share of the labor-intensive aggregate, as well as the changes predicted by the neoclassical factors. Changes of specialization across

---

<sup>18</sup>Production of each HO aggregate is also determined by the relative price of the aggregates, and by the technology used to produce each aggregate. In our framework it turns out that these two effects are empirically equivalent.

aggregates are indeed well explained by the neoclassical factors emphasized here (Rybczynski effect, change in the relative price of the two HO aggregates, technological change biased towards an HO aggregate). As a result, when computing the between-aggregate change, we obtain a good estimate of the change of HI due to the neoclassical factors.

<Figure 3 about here.>

<Figure 4 about here.>

Figures 3 and 4 show that the Rybczynski effect is by far the most important factor explaining the evolution of industrial specialization across HO aggregates. For many countries, the change in the share of the first HO aggregate is close to the observed changes, for both the 1976-1990 period and the 1990-2000 period (i.e. close to the 45-degree line). In both periods, but especially in the first, the results further indicate that, among the neoclassical factors, the Rybczynski effect is hugely dominant as the dots and crosses are nearly coincident in the Figures. It is therefore changes in the countries' capital-labor ratios that explain most of the switch of production from one HO aggregate to the other.

## 4 Conclusion

This paper argues that the HO model provides a possible explanation for the two stages of development documented by Imbs and Wacziarg (2003). The U-curve for industrial concentration observed over an economy's development path may be the direct result of neoclassical factors such as capital accumulation - the traditional Rybczynski effect.

Using a methodology purposely developed to measure the contribution of those neoclassical factors, we find them, especially capital accumulation, to explain between one third and half of the observed changes in industrial concentration.

The decomposition we make presents only a lower bound to the impact of factor accumulation, but it clearly indicates that a substantial fraction of the evolution of industrial diversification patterns remains unexplained by the standard neoclassical factors. Specialization with economies of scale, risk diversification, and other forces are also part of the story. In this paper we proposed a first measure of the total contribution of these non-neoclassical elements. Their respective contributions remain to be determined, however.

## 5 References

- Acemoglu, Daron and Veronica Guerrieri, 2008, "Capital deepening and non-balanced economic growth", *Journal of Political Economy*, 116(3), pp. 467-498.
- Acemoglu, Daron and Fabrizio Zilibotti, 1997, "Was Prometheus unbound by chance? Risk, diversification, and growth", *Journal of Political Economy*, 105(4), pp. 709-751.
- Barro, Robert J. and Jong-Wha Lee, 2001, "International Data on Educational Attainment: Updates and Implications", *Oxford Economic Papers*, 53: 541-563.
- Batista, Catia and Jacques Potin, 2013, "International Specialization and the Return to Capital", working paper, Nova University of Lisbon.
- Deardorff, Alan V., 1974, "A geometry of growth and trade", *Canadian Journal of Economics*, 7(2), pp. 295-306.
- Deardorff, Alan V., 1999, "Patterns of trade and growth across cones", working paper, University of Michigan.
- Dornbusch, Rudiger, Stanley Fischer, and Paul A. Samuelson, 1977, "Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods", *American Economic Review*, 67(5), pp. 823-39.
- Eaton, Jonathan and Samuel Kortum, 2001, "Trade in capital goods", *European Economic Review*, 45(7), pp. 1195-1235.
- Heston, Alan Robert Summers, and Bettina Aten, 2002, Penn World Table Version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP), October.
- Imbs, Jean and Romain Wacziarg, 2003, "Stages of diversification", *American Economic Review*, 93(1), pp. 63-86.
- Koren, Miklós and Silvana Tenreyro, 2007, "Volatility and development", *Quarterly Journal of Economics*, 122(1), pp. 243-287.
- Leamer, Edward E., 1984, *Sources of International Comparative Advantage*, Cambridge, MA: MIT Press.
- Leamer, Edward E., 1987, "Paths of development in the three-factor, n-good general equilib-

rium model", *Journal of Political Economy*, 95 (5), pp. 961-999.

Sachs, Jeffrey D. and Andrew Warner, 1995, "Economic reform and the process of global integration", *Brookings Papers on Economic Activity*, 1, pp. 1-118.

Schott, Peter K., 2003, "One size fits all? Heckscher-Ohlin specialization in global production", *American Economic Review*, 93(2), pp. 686-708.

Trefler, Daniel, 1993, "International factor prices: Leontief was right!", *Journal of Political Economy*, 101, pp. 961-987.

Trefler, Daniel, 1995, "The case of missing trade and other mysteries", *American Economic Review*, 85, pp. 1029-1046.

United Nations Industrial Development Organization, 2005a, *Industrial statistics database, 3-digit ISIC, Revision 2*, Vienna.

United Nations Industrial Development Organization, 2005b, *Industrial statistics database, 3 and 4-digit ISIC, Revision 3*, Vienna.



Coefficient	Labor-intensive HO aggregate	Capital-intensive HO aggregate
$\delta_a^{1981-1985}$	-0.007 (0.057)	-0.007 (0.029)
$\delta_a^{1986-1990}$	0.050 (0.059)	0.032 (0.031)
$\delta_a^{1991-1995}$	-0.011 (-0.059)	-0.040 (0.032)
$\delta_a^{1996-2000}$	0.064 (0.066)	-0.061 (0.037)
$\beta_{1a}$	5.678*** (0.667)	
$\beta_{2a}$		2.237*** (0.087)
$\beta_{3a}$		0.841*** (0.179)
Constraints	$\alpha_{1l} = 0$ $\alpha_{3l} = 0$ $3,000\beta_{1l} = \alpha_{2l}$ $\alpha_{2l} + 18,000\beta_{2l} = 0$	$\alpha_{1k} = 0$ $\alpha_{2k} = 0$ $18,000\beta_{2k} = \alpha_{3k}$

Notes: Estimates of Equation (2) using constrained Non-Linear Least Squares. Country-HO aggregate fixed effects are included in the estimation. Robust standard errors are reported in parentheses. \*\*\* indicates statistical significance at the 1% level. These estimates take the HO aggregate cutoff  $q = 7,500$  to define the two HO aggregates, and the cutoffs  $\tau_1 = 3,000$  and  $\tau_2 = 21,000$  to define the cones.

Table 1: Constrained NLLS estimates for the  $2 \times 2$  model.

Country	$\Delta HI$ x1000	Unadjusted			Adjusted for Composition Effect		
		WE x1000	BE x1000	BE  as %  BE + WE	WE x1000	BE x1000	BE  as %  BE + WE
Australia	-2.1	-13.0	10.8	45.5%	-3.2	1.0	24.9%
Austria	-0.1	-5.7	5.5	49.4%	-1.7	1.5	47.7%
Bolivia	159.6	264.8	-105.2	28.4%	152.5	7.0	4.4%
Canada	1.9	6.1	-4.3	41.0%	1.5	0.4	19.6%
Chile	25.4	1.2	24.2	95.3%	12.9	12.4	49.1%
Hong Kong	-14.6	38.0	-52.6	58.1%	-12.8	-1.9	12.9%
Colombia	5.2	-4.0	9.2	69.6%	6.4	-1.4	17.5%
Cyprus	9.5	12.3	-2.8	18.5%	9.4	-0.1	1.1%
Denmark	2.2	-3.6	5.8	61.8%	0.6	1.7	74.4%
Ecuador	-19.1	-147.5	128.5	46.5%	-8.7	-10.4	54.3%
Egypt	-20.6	-25.7	5.1	16.4%	-24.5	3.6	12.8%
Finland	7.8	2.8	5.0	63.7%	7.4	0.4	5.1%
Greece	6.2	8.5	-2.3	21.5%	8.3	-2.1	19.9%
Guatemala	9.4	6.7	2.7	29.0%	6.7	2.7	28.9%
Hungary	3.9	0.9	3.0	75.8%	3.8	0.0	1.2%
Indonesia	-47.0	-47.0	0.0	0.0%	-47.2	0.0	0.0%
Israel	12.1	-17.4	29.4	62.9%	14.3	-2.0	12.4%
Italy	2.9	-5.2	8.1	60.9%	3.5	-0.5	13.4%
Japan	11.7	11.0	0.7	6.2%	10.9	0.7	6.3%
Jordan	36.7	10.1	26.6	72.6%	19.7	17.0	46.3%
Korea, Rep.	-9.9	-40.9	31.0	43.1%	-0.6	-9.3	93.6%
Malaysia	-11.8	22.6	-34.4	60.4%	-12.0	0.0	0.0%
Netherlands	14.9	7.9	7.0	47.0%	10.5	4.3	29.0%
New Zealand	5.4	-61.9	67.3	52.1%	-2.2	7.5	77.5%
Norway	14.1	1.8	12.3	87.1%	15.4	-1.2	7.3%
Pakistan	-26.2	-25.7	-0.6	2.1%	-25.3	-0.8	3.0%
Panama	9.8	-24.0	33.8	58.5%	-30.9	41.0	57.0%
Philippines	34.6	49.9	-15.3	23.5%	34.5	0.0	0.0%
Portugal	-0.1	-9.3	9.2	49.6%	1.3	-1.5	53.3%
Singapore	44.8	10.1	34.6	77.3%	33.9	10.7	24.0%
Sweden	0.1	-2.7	2.8	51.2%	-47.6	-8.1	14.5%
Turkey	15.3	-27.6	42.9	60.8%	-6.4	21.6	77.3%
UK	-3.1	-6.1	2.9	32.7%	-0.1	-3.0	96.5%
Tanzania	22.2	22.2	0.0	0.0%	22.3	0.0	0.0%
USA	-3.0	-20.7	17.7	46.1%	-1.7	-1.2	41.6%
Venezuela	-7.3	-14.0	6.7	32.3%	-18.6	11.2	37.5%
Average (N=36)	8.1	-1.2	9.2	46.1%	4.0	2.5	29.7%

Table 2: Decomposition of  $\Delta HI$ , 1976-1980 to 1986-1990.

Country	$\Delta HI$ x1000	Unadjusted			Adjusted for Composition Effect		
		WE x1000	BE x1000	BE  as %  BE + WE	WE x1000	BE x1000	BE  as %  BE + WE
Austria	-1.1	-1.1	0.0	0.0%	-1.1	0.0	0.0%
Bangladesh	-48.9	-48.9	0.0	0.0%	-48.7	0.0	0.0%
Bolivia	-46.2	-214.2	168.0	44.0%	-180.8	134.7	42.7%
Chile	4.7	5.4	-0.7	12.0%	-40.8	45.4	52.7%
Hong Kong	-11.5	-87.8	76.3	46.5%	-1.6	-9.8	85.7%
Colombia	18.0	12.4	5.6	31.3%	13.0	5.1	28.1%
Cyprus	39.1	-1.5	40.7	96.3%	38.4	1.0	2.4%
Ecuador	135.6	76.8	58.9	43.4%	169.5	-33.7	16.6%
Egypt	10.7	-17.8	28.4	61.5%	3.7	7.2	66.4%
El Salvador	-44.4	-44.4	0.0	0.0%	-44.6	0.0	0.0%
Finland	15.8	12.5	3.3	20.6%	27.0	-11.1	29.2%
Greece	20.9	26.8	-5.9	18.1%	38.3	-17.6	31.5%
Hungary	23.6	13.4	10.2	43.3%	23.7	0.0	0.1%
India	12.0	45.5	-33.5	42.4%	11.9	0.0	0.0%
Indonesia	-28.6	7.8	-36.4	82.4%	-28.5	0.0	0.0%
Israel	-3.8	-2.5	-1.4	35.9%	12.7	-16.7	56.7%
Italy	17.6	17.6	0.0	0.0%	17.6	0.0	0.0%
Japan	4.7	3.5	1.2	26.5%	12.9	-8.1	38.6%
Jordan	-10.2	117.3	-127.5	52.1%	34.0	-44.4	56.6%
Korea, Rep.	20.3	7.4	12.9	63.5%	35.5	-15.2	30.0%
Malaysia	40.9	-71.5	112.4	61.1%	45.1	-4.2	8.6%
Mexico	-2.9	18.2	-21.1	53.7%	0.3	-3.1	92.1%
Netherlands	2.1	2.1	0.0	0.0%	2.1	0.0	0.0%
Nigeria	16.0	16.0	0.0	0.0%	16.1	0.0	0.0%
Norway	-9.5	-5.3	-4.2	43.8%	-3.1	-6.4	67.5%
Pakistan	-8.6	-116.4	107.8	48.1%	-20.1	11.4	36.2%
Panama	66.1	195.7	-129.7	39.9%	65.9	0.0	0.0%
Peru	27.0	40.1	-13.2	24.7%	27.1	0.0	0.0%
Philippines	-49.4	-25.2	-24.2	49.0%	-49.9	0.5	1.0%
Poland	-13.7	-10.5	-3.2	23.2%	-13.7	0.0	0.0%
Portugal	-13.7	-17.8	4.1	18.8%	-13.3	-0.4	2.7%
Singapore	40.3	26.2	14.1	34.9%	77.6	-37.0	32.3%
Spain	-1.8	2.3	-4.1	63.8%	14.3	-16.0	52.9%
Sri Lanka	-88.5	-88.5	0.0	0.0%	-88.7	0.0	0.0%
Turkey	-17.0	-46.7	29.7	38.9%	-4.3	-12.6	74.4%
UK	0.5	-0.6	1.1	65.4%	6.5	-6.0	48.1%
USA	3.3	0.8	2.5	75.2%	21.9	-18.6	45.9%
Venezuela	48.8	-155.1	203.9	56.8%	48.0	0.9	1.9%
Average (N=38)	5.3	-7.0	12.3	36.6%	6.2	-1.0	26.9%

Table 3: Decomposition of  $\Delta HI$ , 1986-1990 to 1996-2000.

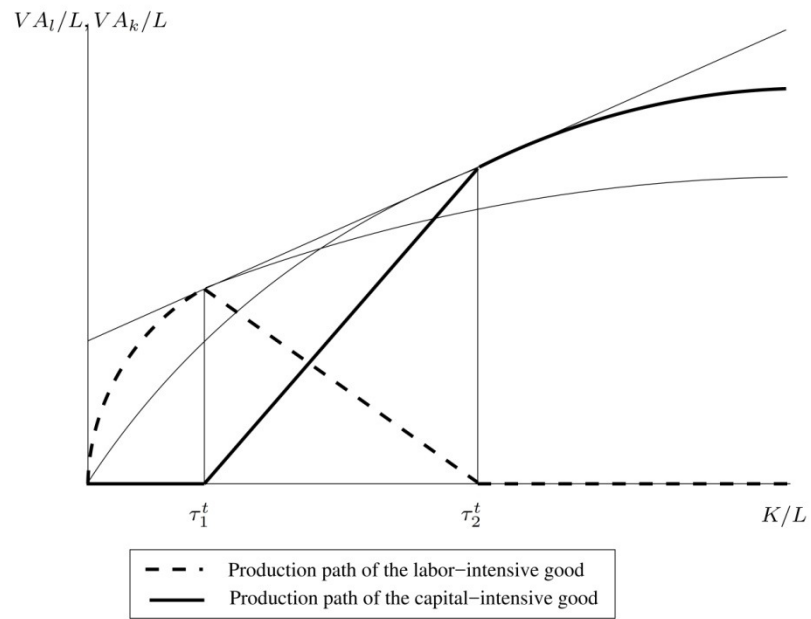


Figure 1: Production Patterns in the 2x2 HO Model.

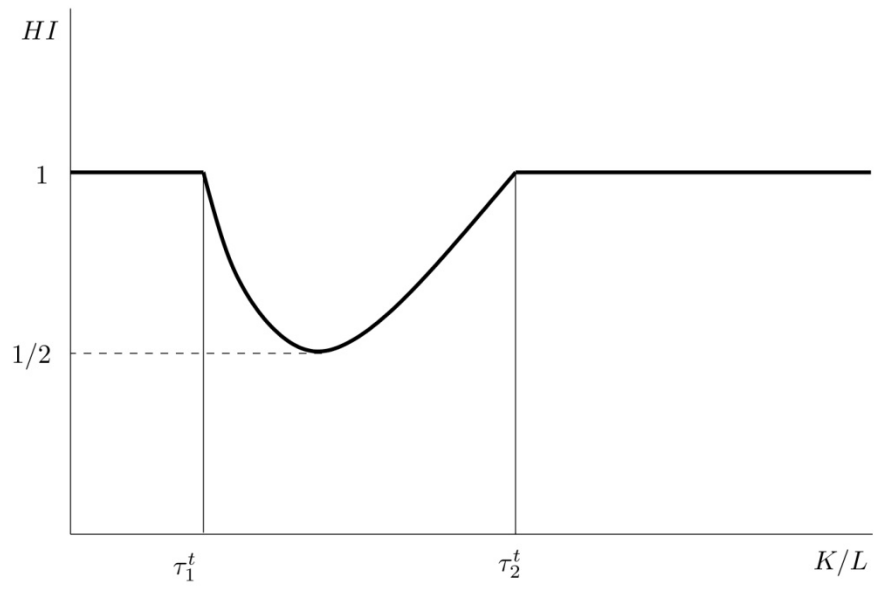


Figure 2: The Herfindahl index in the  $2 \times 2$  HO model.

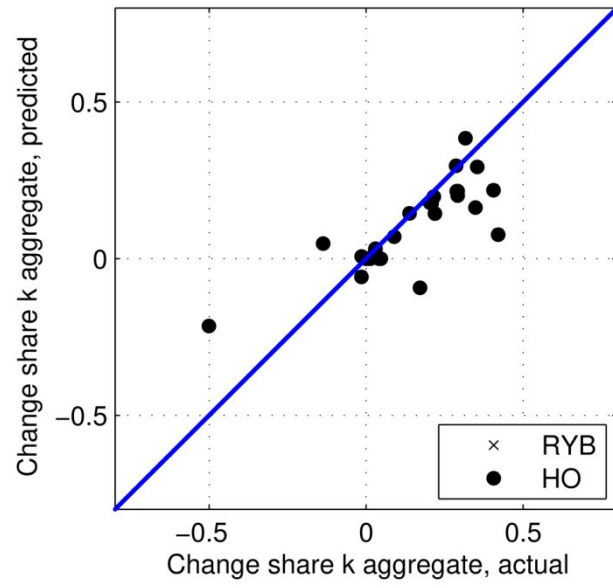


Figure 3: Changes in the share of the capital-intensive HO aggregate, actual, attributed to HO factors, and attributed to the Rybczynski effect only, (1976-1980) - (1986-1990).

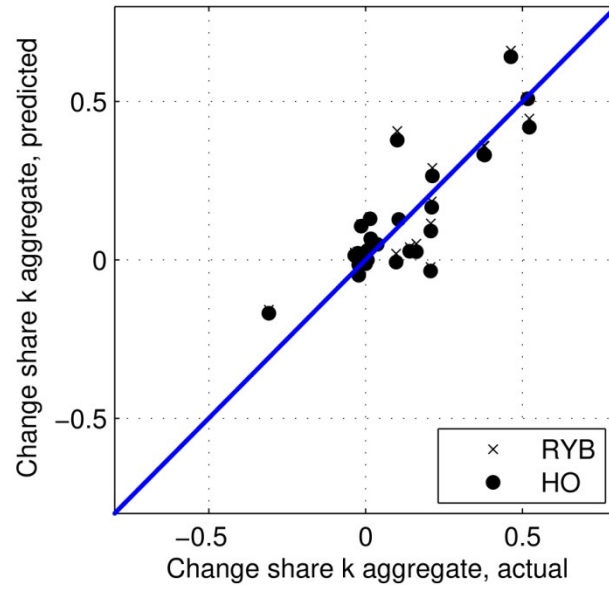


Figure 4: Changes in the share of the capital-intensive HO aggregate, actual, attributed to HO factors, and attributed to the Rybczynski effect only, (1986-1990) - (1996-2000).