# A Compositional Statistical Analysis of Capital Stock per Worker*

Juan Manuel Larrosa†
*CONICET – Universidad Nacional del Sur*
*jlarrosa@criba.edu.ar*

*Abstract*

Most of economic literature has presented its analysis under the assumption of homogeneous capital stock composition. However, capital composition differs across countries. What has been the pattern of capital composition associated with World economies? We make an exploratory statistical analysis based on the Aitchinson logratio transformations and the related tools for visualizing and measuring statistical estimators of association among the components. As initial findings could be cited that:
(1) It is observed a clear correlation in terms of capital stock participation between two building-industry-related components,
(2) Manufacturing behaves differently, especially durable goods sector.
(3) There's differences among subsamples

*Resumen*

Gran parte de la literatura económica analiza al capital físico como un stock homogéneo. Sin embargo, la composición del capital difiere entre países. ¿Cuál ha sido el patrón de composición de capital asociado a las economías del mundo? Realizamos un análisis estadístico exploratorio basados en las transformaciones logcocientes de Aitchinson y en herramientas para visualización y medición de estimadores de asociación entre componentes. Inicialmente se ha hallado:
(1) Existe clara correlación en la participación del capital entre dos sectores relacionados con la industria de la construcción.
(2) La industria manufacturera se comporta de manera diferente
(3) Existen diferencias entre submuestras.

*JEL Classification: C82, E22*

## 1. Introduction

While physical capital stock represents a crucial factor in the economic process, less is known about the joint behavior of capital components. This paper tries to show first results about how the composition of capital has performed during the 1965-1990 period for a

* Acknowledgment:
I would like to thanks V. Pawlowsky-Glahn for comments and corrections to a very early version of this paper. All errors are mine.

† *Contact Address: Departmento de Economía, Universidad Nacional del Sur, San Juan y 12 de Octubre, Planta Baja, Gabinete 5, Bahía Blanca 8000, Argentina*.

heterogeneous sample of countries. We used statistical tools for visualizing patterns in the data sample as well as recent economic evidence to show some possible explanations.

Given that we are asking about capital components, we should use data that reflects its composition and variability. We used compositional data that consists of positive valued vectors summing to a unit (hundred per cent), or in general to some fixed constant for all vectors. Examples of this kind of data in economics are many, including household budget shares, aggregate output composition, shareholder's portfolio composition, etc. Lack of statistical independence condemns this type of data for using typical statistical inference methods. It follows that some transformation, if it exists, has to be applied before analysis. Fortunately in our case it exists, and allows for the use of almost full multivariate analysis procedures. Our goal is to find patterns in the capital per worker composition looking for answers about how these components have behaved. This behavior should be interpreted as the struggle among economic sectors for capital allocation. We found a clear correlated behavior in building sectors and a fuzzier correlated behavior in machinery and equipment sectors in the full sample, and in two of three subsamples. Interestingly, the subsample with an odd performance refers to a group of high rate of growth countries.

The paper is organized as follows. Section 2 summarizes recent literature on physical capital investment behavior. Section 3 describes the statistical theory that supports the analysis. Section 4 presents the sample and subsample analysis results and section 5 ends with preliminary conclusions and discussion.

## 2. Literature on physical capital patterns

Several works have emphasized the importance of specific capital investment as requirements for growth. Since De Long and Summers (1990) shaded light to the roll of equipment investment in the growth process for a sample of countries during the period 1960-1985, many other research works supported this finding in the broad sense (for example, Temple and Voth, 1998.) At the same time, Jones (1994) investigated how affected is growth by distortions in capital relative price. Working with some of the same variables of this paper, Jones found that higher relative price of capital (through taxes or tariffs on importing) resents growth. Explicitly, he found negative correlation between all capital subaggregate components relative prices and annual growth rate per capita. In a more theoretical framework, Jovanic and Rob (1997) used a modified Solow growth scheme for modeling the fact that machinery is more expensive in less developed countries. However, the most insightful research into particular components of capital stock of the economies

could be found in a research paper series supported by the World Bank that will be following summarized.

Canning (2000) develops a panel data production function estimation that includes as infrastructure variables: miles of roads, electricity generating capacity, and telephones per workers. He found that only the variable telephones per worker is statistically significant in the sample, suggesting that this variable generates more externalities in the economy than the first two. Ingram and Liu (1997) estimated the influence of economic variables in a wide range of equipment and transportation variables in a heterogeneous sample of countries and cities. Their work shed light on the pros and against the high level of motorization in big cities and the externality that this provokes in prices of land, congestion, and pollution. As they recalled in another related work (Ingram and Liu, 1999) in the past 15 years the World stock of vehicles grew up in about 60%, because of lower production costs and a higher income in less developed countries. This way it could be expected a significant participation of transportation capital in the total stock of capital. Again, the question remains of whether this increment has been done by taken participation to other class of capital. Randolph *et al.* (1996) found a set of variables that correlates positively with investment in infrastructure related to transportation and communication sector. It could be mentioned the urbanization level, foreign sector size, population density and funding mechanism.

A crucial feature related to infrastructure investment is how these projects are funding and financing. Klingebiel and Ruster (2000) summarize that most of governments induce private sector to invest in infrastructure through soft lending, guarantees, and grants with a wide variety of results. This inducement process has had very different results depending on the institutional framework implemented and the specific financed project, but this shows how infrastructure market is an active one, not only wrapped around the government hand. But government investment has a crucial roll in this aspect. Reinikka and Svensson (1999) studied the cases of less developed countries where in some cases they assured that government investment in infrastructure is even more important than macroeconomic stability in the private sector investment decision process. Infrastructure provides through costs reductions and linkages positive externalities to economy as a whole.

At the same time, the building sector shows itself as a highly expansive one in whether developed and undeveloped countries. Housing is upraising in the developed countries because people are moving from downtown to the suburbia. This behavior is robust to different kinds of shocks like those studied by Glaeser and Gyourko (2001) for the American case. New construction is enhanced by lower land prices and lower mortgage rates in developed countries. In the other hand, in less developed countries housing represents a substantial part of the capital stock because its less industrialized profile.

Nevertheless physical capital components are markedly complementaries. The building of a dam requires not only of concrete and rolling stones but also of road infrastructure and housing for the workers. Canning and Bennathan (2001) studied the social rate of return of generating capacity of electricity and paved roads projects and showed that both kinds of projects reflects higher than average rates of returns when considered simultaneously. In isolation, both kinds of projects reflect lower than social rates of return. That's because when they considered investments' potential benefits against its construction cost, complementarities emerge in a crossed way. This supports the idea of considering a mix of capital components when analyzing infrastructure investment, a key issue in the interpretation of the present work that we'll considerer as the complementarity approach.

Another kind of physical capital is inventories. Guasch and Kogan (2001) survey the inventories statistics of a sample of countries and found that less developed countries have three times more inventories stocks than developed countries. The problem associated with keeping high inventories is usually lack of efficiency in the industry structure, transforming this inefficiency into tangible results with lower benefits (lost transactions, delays in deliveries, high amount of immobilized capital). Again, the low rate of investment in new depots or warehouses and the small market size does not help much in solving the problem in developing countries. They found that inventories levels are correlated negatively with GDP per capita and a dummy variable that counts for infrastructure quality.

Table 1 concisely reports main findings of the literature review and focuses in the main variables related to physical components analyzed by each research paper.

**Table 1. Summary of references**

| Author/s | Capital Component | Results (type of data or analysis) |
|---|---|---|
| De Long and Summers (1990) | Equipment and machinery investment | Positive correlation between growth rate and equipment and machinery investment (country data). |
| Temple and Voth (1998) | Equipment and machinery | Positive correlation between growth rate and equipment and machinery investment (country data) |
| Jovanovic and Rob (1997) | Equipment and machinery | Machinery is more expensive in less developed countries (country data) |
| Hall and Jones (1998) | Physical Capital | Positive relation between social infrastructure (as defined by the authors) and capital intensity (measured as total capital stock per worker) (country data). |
| Jones (1994) | Physical capital and components relative price | Negative correlation between capital component relative prices and growth (country data) |
| Canning (2000) | Non-residential construction and transportation equipment | A variable telephone per worker is statistically significant in explaining countries' aggregate output (country data). |

| | | |
|---|---|---|
| Ingram and Liu (1997) | Durable goods and transportation equipment | Geographic and economic (country and urban) variables significantly correlated with motorization and transportation variables. |
| Ingram and Liu (1998) | Durable goods and transportation equipment | Environment and economic (country and urban) variables significantly correlated with motorization and transportation variables. |
| Randolph *et al.* (1996) | Transportation equipment | Social, economic and institutional variables significantly correlated with public investment in transportation infrastructure (country data) |
| Klingebiel and Ruster (2000) | Infrastructure investment | Importance of private sector participation in infrastructure provision (case studies) |
| Reinikka and Svensson (1999) | Infrastructure investment | Importance of government infrastructure investment in private sector investment expectations (firm data) |
| Glaeser and Gyourko (2001) | Residential building | Several economic, social, and infrastructure variables explained significantly housing rates (urban data) |
| Canning and Bennathan (2001) | Non-residential construction | Importance of considering mixes capital components in infrastructure analysis –for covering complementarities and externalities effects (country data). |
| Guasch and Kogan (2001) | Equipment investment (inventories) | Negative correlation between inventories level and GDP per capita and infrastructure quality dummy (country data) |

An interesting question that remains unanswered is the potential displacement of a class of capital by another during the economic process. Equipment investment could displace durables goods in the total capital participation? How are complementarities present in capital composition? We will see that some clues for these questions could be obtained by using capital compositional data and specific statistical techniques and procedures.

## 2. Statistical Model and Techniques

We worked with compositional data then we briefly introduce definitions and analytic techniques for the processing of this specific kind of data. Compositional data refers to vectors of data that represent proportions of a whole. Assume a vector $x$ with non-negative elements $x_1,\ldots,x_D$. If we normalize $z_i = x_i / X$ where $i \in (1,\ldots,D)$ and $X = \sum_{i=1}^{D} x_i$ then we have that

$$z_1 + z_2 + \cdots + z_D \equiv 1 \qquad (1.1)$$

The problem that arises with this data structure when doing statistical inference is that inference is subject to the unit-sum constraint (1.1). Pearson (1897) gives the first warning about the difficulties in the statistical inference process that can be found by modeling this kind of data. He noticed that when attempting to estimate correlation with indices it was likely to emerge spurious correlation. The source of the problem could be found in that numerator

and denominator stand in the variable at the same time, $z_i$ in (1.1), then independence is violated for statistical inference.

Hopefully Aitchinson (1986) proves how to deal with this problem by adequate transformations in the data that left them ready for multivariate statistical inference[1]. In fact, these transformations have highly desirable properties as scale invariance, subcompositional coherence and perturbation invariance in the simplex. Scale invariance refers to the propriety of the transformation not to alter the composition in terms of its components shares and distances after the transformations have been taken place. Subcompositional coherence refers to correspondence in the product-moment correlation between raw components as a measure of dependence after subcompositions has been created. Finally, perturbation to a compositional vector should be restored to original data after the inverse of the initial perturbation has applied to the transformed vector.

By sake of clarity, we must define usual operations applied to compositional data: the perturbation and closure operation. Perturbation of one composition *x* by another composition *y* refers to the operation

$$x, y \in S_c^d \Rightarrow x \circ y = \mathbb{C}\left(x_1 y_1, x_2 y_2, \ldots, x_d y_d\right) \in S_c^d,$$

which is termed a *perturbation* with the original composition *x* being operated on by the *perturbing* vector *y* to form a *perturbed* composition $x \circ y$. $\mathbb{C}(.)$ refers to the closure operation, defined for any vector $z = \left(z_1, z_2, \ldots, z_d\right) \in \mathbb{R}_+^d$ by

$$\mathbb{C}(z) = \left( \frac{z_1}{\sum_{i=1}^d z_i}, \frac{z_2}{\sum_{i=1}^d z_i}, \ldots, \frac{z_d}{\sum_{i=1}^d z_i} \right)$$

Center (also called baricenter) or the geometric mean closure of an *N* size sample is defined by $g_m = \mathbb{C}\left(g_1, g_2, \ldots, g_d\right)$, with $g_i = \left(\prod_{n=1}^N x_n\right)^{1/N}, i = 1,2,\ldots,d$ and represents accurately the sample central trend. When we perturbed a compositional dataset by the baricenter inverse, we centered the data allowing for better visualization of data structure.

Another two important concepts have to be defined: subcomposition and amalgamation. *Subcompositions* are obtained when we take two or more components of the composition and then we closed them. We analyze the subcomposition as a composition in itself when this could be interesting for the purpose of the research. More formally, the subcomposition based on parts (1,2,…,*C*) of a *D*-part composition ($x_1$, $x_2$,…, *D*), where *C*<*D*, is the (1,2,…C)-subcomposition ($s_1$,$s_2$,…,$s_C$) defined by

$$\left(s_1, s_2, \ldots, s_C\right) = \frac{\left(x_1, \ldots, x_C\right)}{\left(x_1 + \cdots + x_C\right)},$$

the *closure* operation. Finally, another useful tool for compositional analysis is the

*amalgamation* procedure. Following Aitchinson (1986, pp. 37) we state that when if parts of a *D*-parts composition are separated into *C* ($\leq D$) mutually exclusive and exhaustive subsets and the components within each subset are added together, the resulting *C*-part composition is termed an *amalgamation*. Then, amalgamation is when we select reasonably components and add them for obtaining new ones. Following we briefly summarize the transformations we are going to use for the data analysis since it is not usually observed in applied economic analysis.

When we use compositional data we recognize that magnitude is irrelevant, because we analyze the information on relative proportions of the registered components. Therefore, any transformation of a compositional data set has to be invariant by the group to scale changes, i.e., it has to be likely expressed in terms of ratios of the composition components. Aitchinson (1986) defines the two transformations we are going to use for analytic purposes: First, the centered logratio transformation (*clr*) is the bijective application between $x \in S_c^d$ to $z \in \mathbb{R}^d$ defined by

$$clr(x) = \left( \ln \frac{x_1}{g(x)}, \ln \frac{x_2}{g(x)}, \ldots, \ln \frac{x_d}{g(x)} \right) = (z_1, z_2, \ldots, z_d), \tag{1.2}$$

with $g(x) = \left( \prod_{i=1}^{d} x_i \right)^{1/d}$ as the geometric mean of the composition. The inverse of the transformation in this case is $clr^{-1}(z) = \mathbb{C}(\exp(z_1), \exp(z_2), \ldots, \exp(z_d)) = (x_1, x_2, \ldots, x_d)$, where $\mathbb{C}(.)$ represents the *closure* operation. Notice that in *clr* transformation, geometric mean is estimated by using data matrix rows (observations) while in the definition of the center of observations set (ternary diagram center), geometric mean is calculated by columns (variables).

Second, the additive logratio transformation (*alr*) is the bijective application from $x \in S_c^d$ to $y \in \mathbb{R}^d$ defined by

$$alr(x) = \left( \ln \frac{x_1}{x_d}, \ln \frac{x_2}{x_d}, \ldots, \ln \frac{x_{d-1}}{x_d} \right) = (y_1, y_2, \ldots, y_d), \tag{1.3}$$

the inverse of this transformation is called the additive-logistic generalized transformation defined by $agl(y) = \mathbb{C}(\exp(y_1), \exp(y_2), \ldots, \exp(y_{d-1}), 1) = (x_1, x_2, \ldots, x_d)$.[2]

Once we have obtained the transformed data, some procedures would help us to understand the joint variability of the analyzed variables. Such instruments are prediction and confidence regions (both analogous to prediction and confidence intervals). For these tools to be calculated it must be used *alr* transformation in the data. In the case of prediction regions we estimate the isoprobability ellipses from the corresponding multivariate normal

distribution in the real space $\mathbb{R}^{d-1}$ and then we applied *agl* to these ellipses. Formally, (see Aitchinson [1986], pp. 174-176) a predictive region of content *c* for a *D*-part composition **x** based on the experience of compositional data set X is given by

$$\left\{x : q\left[alr\left(x\right)\right] \le q\right\}$$

where $q\left(y\right) = \left(1+N^{-1}\right)^{-1}\left(y-\hat{\mu}\right)'\hat{\Sigma}^{-1}\left(y-\hat{\mu}\right)$ and $q$ is predetermined by Fisher distribution function statistic

$$\mathrm{F}_{q/(q+n)}\left\{\tfrac{1}{2}d,\tfrac{1}{2}\left(n-d+1\right)\right\} = c$$

In the case of confidence regions over the group baricenter we rely on the transformed *alr* composition multivariate normality hypothesis[3]. Formally, assume a random composition $x \in S_c^d$ and assume that it follows a normal logistic additive distribution. Then $y$ = *alr*($x$) follows a ($d$ –1)-dimensional normal multivariate distribution and for a sample of size N we have a predictive region defined by

$$\frac{N\left(N-d-1\right)}{\left(d-1\right)\left(N-1\right)}\left(\bar{y}-\mu\right)\hat{\Sigma}^{-1}\left(\bar{y}-\mu\right),$$

where $\mu = E\left[x\right]$ and $\hat{\Sigma}$ is the covariance matrix $\Sigma$ maximum likelihood estimator. This is an estimator that follows a Fisher distribution F with ($d$ –1, $n$ – $d$ –1) degrees of freedom. Then, for given $\alpha$, tables give $\kappa$ such that

$$1-\alpha = \mathrm{P}\left[\frac{N\left(N-d-1\right)}{\left(d-1\right)\left(N-1\right)}\left(\bar{y}-\mu\right)\hat{\Sigma}^{-1}\left(\bar{y}-\mu\right)' \le \kappa\right]$$

$$= P\left[\frac{N\left(N-d-1\right)}{\left(d-1\right)\left(N-1\right)}\left(\bar{y}-\mu\right)\hat{\Sigma}^{-1}\left(\bar{y}-\mu\right)' \le \frac{N\left(N-d-1\right)}{\left(d-1\right)\left(N-1\right)}\kappa\right]$$

Then, $\left(\bar{y}-\mu\right)\hat{\Sigma}^{-1}\left(\bar{y}-\mu\right)'$ = constant defines an ellipse centered at $\bar{y}$ in $\tilde{\mathrm{N}}^2$, which can be plotted by finding the pairs of values $\mu_1$ and $\mu_2$ which form the vector $\mu$ and satisfy the equation. Consequently, $\xi$=*agl*($\mu$) defines a confidence region around the center in the ternary diagram or simplex. Now we concisely report the statistical tool we are going to use for the transformed data analysis.

Principal components analysis (PCA) refers to the analysis of the invariant properties of the covariance matrix of a data sample that allows for reducing dimensionality in the data structure (see Aitchinson 1986, Section 8.3, for the compositional case). The basic idea in PCA is to find the components $s_1,s_2,...,s_n$ so that they explain the maximum amount of variance possible by $n$ linearly transformed components. Data covariance matrix it's the main source of information and from it we estimate its eigenvalues. Eigenvalues are denominated

principal components and sorted from highest to lowest. Thus the first principal component is the projection on the direction in which the variance of the projection is maximized.

As we mentioned before, the basic goal in PCA is to reduce the dimension of the data. Indeed, it can be proven that the representation given by PCA is an optimal linear dimension reduction technique in the mean-square sense. Such a reduction in dimension has important benefits. First, the computational overhead of the subsequent processing stages is reduced. Second, noise may be reduced, as the data not contained in the *n* first components may be mostly due to noise. Third, a projection into a subspace of a very low dimension, for example two, is useful for visualizing the data, in our case we will try to reduce in a way that can be represented into the simplex. Note that often it is not necessary to use the *n* principal components themselves, since any other orthonormal basis of the subspace spanned by the principal components has the same data compression or denoising capabilities.

The information that we are going to use from PCA is mainly visual: the biplot. This is a powerful tool for exploratory analysis. It shows an axis diagram where the axis intersection (*origin O*) represents the center of the sample, for each of the *d* components there's a *vertex* $v_i$ in the position $h_i$, $i = 1,2,\ldots,d$, and for each of the *N* cases, there's a *marker* $\alpha_n$ in $g_n$, $n = 1,2,...,N$.

The union of *O* to vertex $v_i$ is denominated radius $\overline{Ov_i}$ or *ray*, and the union of two vertexes $v_i$ and $v_j$ is named *link* $\overline{v_i v_j}$. They represent statistical information because

$$\left|\overline{v_i v_j}\right|^2 \approx \text{var}\left[\ln\frac{x_i}{x_j}\right], \qquad \left|\overline{Ov_i}\right|^2 \approx \text{var}\left[\ln\frac{x_i}{g(x)}\right],$$

where $|\cdot|$ represent the length of the segment. At the same time,

$$\cos\left(v_i O v_j\right) \approx corr\left[\ln\frac{x_i}{g(x)}, \ln\frac{x_j}{g(x)}\right],$$

and if *links* $\overline{v_i v_j}$ and $\overline{v_h v_\ell}$ intersect in *M*, then

$$\cos\left(v_i M v_h\right) \approx corr\left[\ln\frac{x_i}{x_j}, \ln\frac{x_h}{x_\ell}\right],$$

When two links lie approximately in straight angle, then $\cos\left(v_i M v_h\right) \approx 0$ implying that correlation between the two ratios is almost null. This is helpful information for the search of possibly independents subcompositions.

Another interesting aspect is observed when vertexes $v_i$ and $v_j$ coincide, or almost, then $\left|\overline{v_i v_j}\right| \approx 0$, resulting that $\text{var}\left[\ln\left(x_i/x_j\right)\right] \approx 0$, i.e., $x_i/x_j \approx$ constant. If we represent these

two components in any other section of the ternary diagram we should see that their relationship is linear. Two more interesting features are: first, when a subset of vertexes is collinear then the associated subcomposition has a biplot that is one-dimensional, that is, it has unidimensional variability. Finally, biplots stand invariant whether we perturb the data set or not. That's because perturbations don't change the data covariance structure. Graphically, *links* represent column information while dots individual information. Angles between lines represent the correlation between columns (variables)[4].

## 3. Data structure and analysis

We begin this section by defining the variables we are going to use. Data were extracted from Penn World Table 5.6 and corresponds to KDUR, KOTHR, KNRES, KRES, and KTRAN series for the 1965-1990-time period. A brief description of these is published in Table 2. Series were selected according to the following criteria: a) countries were included only if they had full data series, and b) countries with zeros in any series were discarded, although there exists procedures to take this case under reasonably control for statistical inference (see Martín-Fernandez *et al*, 2000, and Fry *et al*., 2000) we decided not to deal with this particular problem because of the original exploratory goal.

**Table 2. Code and description of variables**

| Index | Code | Description |
|---|---|---|
| 1 | **KDUR** | Percentage of capital per worker allocated in durable production assets (machinery and equipment). |
| 2 | **KOTHR** | Percentage of capital per worker allocated in other buildings. |
| 3 | **KNRES** | Percentage of capital per worker allocated in non-residential building. |
| 4 | **KRES** | Percentage of capital per worker allocated in residential building. |
| 5 | **KTRAN** | Percentage of capital per worker allocated in transportation equipment. |

Series were presented initially as percentages of the capital stock per worker in 1985 international prices. This fact made that total sum of components was different from unit in different periods. We proceeded by bounding the composition y closing each compositional vector year by year. So we've got, for each year, the participation of each compositional vector in the hundred percent of each economy's capital stock per worker. Then we calculated the geometric mean of each vector for all the analysis period and closed it again because geometric mean of variables was less than the total explanation. This way we obtain the average participation of each compositional vector for the time span of the sample. In Appendix 1 raw data used in this work is published together with the country list.

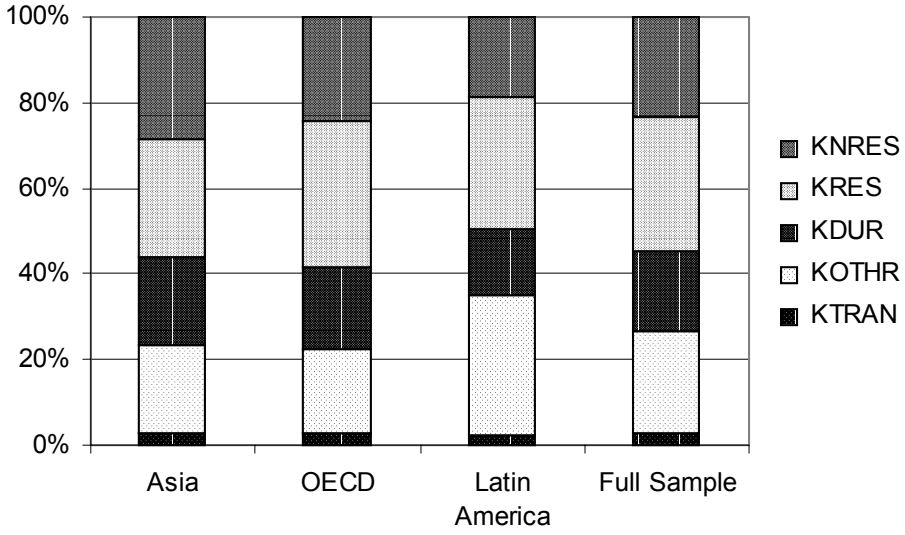
Once we obtain the final raw data block, we proceed to transform them with the centered logratio transformation *clr*. This imply that we should apply (1.2) defined by

$$clr(x) = \left( \ln \frac{x_1}{g(x)}, \ln \frac{x_2}{g(x)}, \ln \frac{x_3}{g(x)}, \ln \frac{x_4}{g(x)}, \ln \frac{x_5}{g(x)} \right)$$

with $g(x) = (x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5)^{1/5}$ and 1,…,5 represents the index for the components in Table 1. Given that this transformation preserves the distance among data results more useful for multivariate statistical analysis.

Full sample raw data descriptive statistics is published in Table 4 in the Appendix at the end of this paper. As we can see KTRAN seems to be the more volatile variable, while KRES is the more stable compositional variable of the full sample. Figure 1 shows stacked bars for the full sample and subsamples. These will be described later.

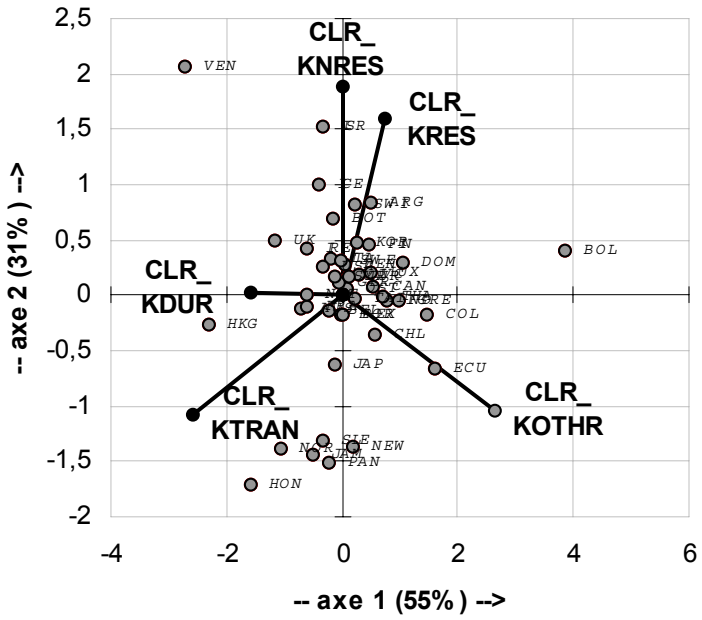
**Figure 1. Comparative raw data for geographic and economic regions**




### 3.1 Full Sample Analysis

*Clr*-transformed data allows us to full utilization of multivariate tests (transformed variables are denoted with a CLR_ prefix). PCA using the covariance matrix was calculated on the five compositional vectors and the biplot is published in Figure 2 (total explained variability is in parenthesis). Table 8 in the Appendix 2 describes the statistical results of these estimations. There, it can be checked out the magnitude and sign of the relationship illustrated in Figure 2.

**Figure 2. Biplot on the first two principal components (84%) – Full sample**




Near coincident vertices are observed in CLR_KNRES and CLR_KRES while CLR_KTRAN shows collinearity with CLR_KRES. CLR_KDUR shows a non-correlated behavior with CLR_KNRES, and an almost (perpendicular) non-correlated behavior with CLR_KRES and, at a lower extent, to CLR_KOTHR. Lastly, CLR_KOTHR shows a particular scarce correlated behavior with any of the aforementioned variables.

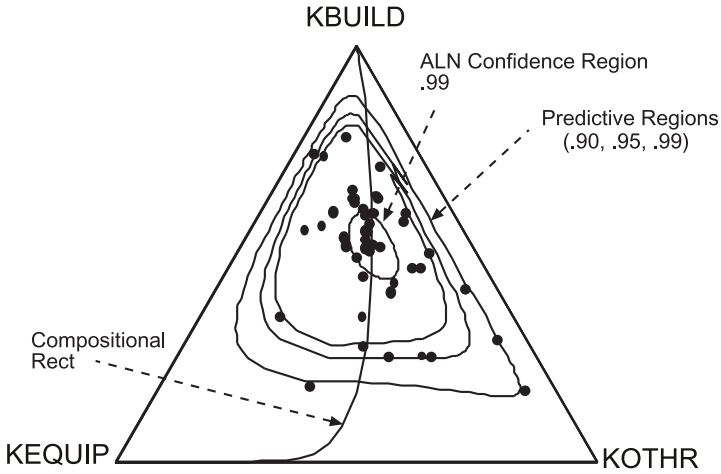
Because of the variable definition it could be anticipated that there's different behavior in the composition of the capital stock whether an economy builds or manufactures machinery or equipment. Residential and non-residential buildings behave in a correlated way, while transportation equipment seems to be negative correlated with these. One can argue that machinery manufacturing (transportation in one hand, and vehicles and durable goods in the other) behaves differently from building (residential and non-residential). It could be reasonably to amalgamate (join and close) both series under a functionality sorting that could be categorized as capital allocated in building industries and capital allocated in equipment (in the broad sense). Then we decided to amalgamate KNRES and KRES into a new variable called KBUILD and the same has been done with KTRAN and KDUR into KEQUIP. KOTHR remains the same because its particular definition and observed behavior (uncorrelated). These transformations could be clearly summarized observing Table 3.

**Table 3. Amalgamation and labeling of new variables**

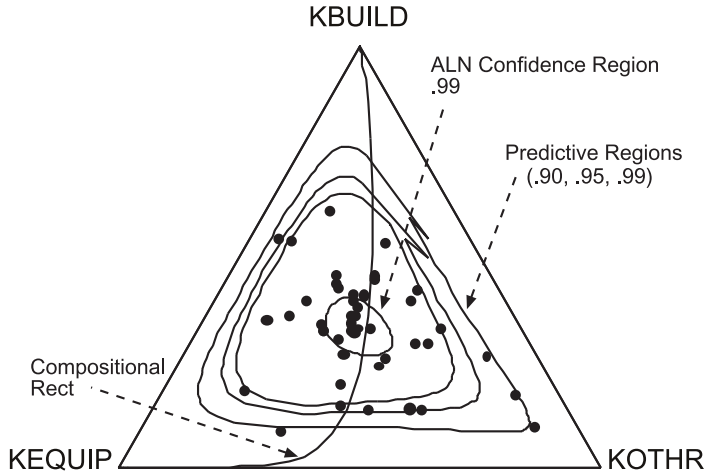
| Raw Variable | Amalgamated Variable |
|---|---|
| KNRES<br>KRES | KBUILD |
| KDUR<br>KTRAN | KEQUIP |
| KOTHR | KOTHR |

Now we have reduced the dimensionality of the original dataset and we deal with three compositional vectors that can be visualized through the ternary diagram or simplex. This is shown in Figure 3a. Some data are accumulated near the upper center of the graph, with several isolated points. We perturbed this data by centering them and we obtain the Figure 3b. In both figures we included the confidence region for the baricenter and the predictive regions for the whole sample (at .99, .95, and .90% of significance level). Likewise we included a compositional rect that allow us to observe a linear relationship between KEQUIP and KBUILD that we foresaw at the biplot in Figure 2.

**Figure 3. Amalgamated data on the simplex (Full Sample)**




*Raw Data* (**3.a)**




*Centered* (**3.b)**

Most data lied on the .99 predictive region but several points lied on the .95 and .90 regions also, and some of them seem to be outliers. This could indicate that several statistical populations are located in the sample. Later, we will try to cluster this points but with economic and geographic information.

KEQUIP and KBUILD relationship could be remarked by picturing the *alr*-plot on these two components in relation to KOTHR. We transformed these ratios with (1.3) and plotted them scattered. This is shown in Figure 4 (linear trend included).

**Figure 4. Alr-plot on KEQUIP and KBUILD (Full Sample)**

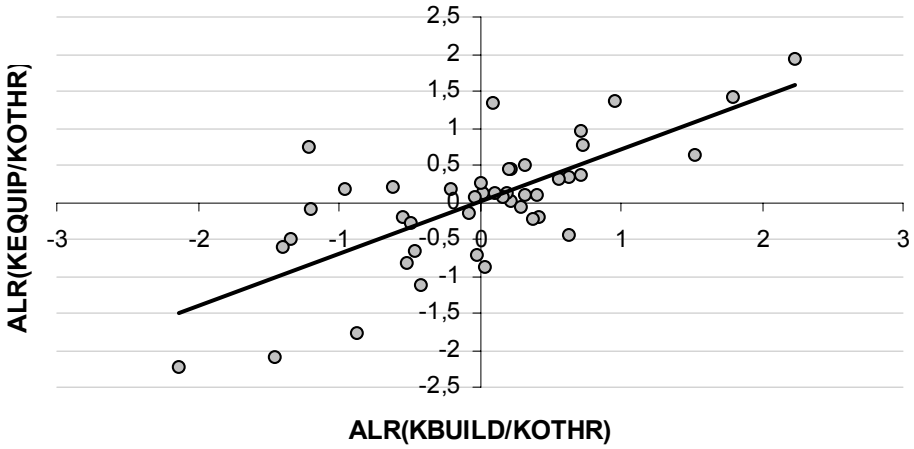


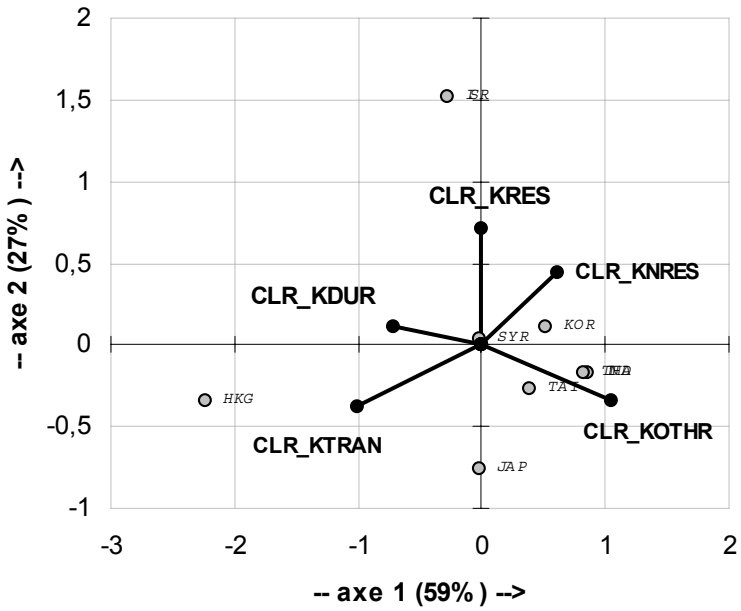

Figure 4 describes a positive relation between both ratios. It seems that equipment investment and building shares move together, as pointed out by Canning and Bennathan (2000)'s complementarities approach.

### 3.2 Subsample Analysis

Now we follow the same procedure applied in Section 3.1 to the geographic and economic subsamples. Figure 3.a and 3.b showed the potential existence of different populations into the sample. We define subpopulations in terms of economic and geographic reasons[5]. Due to the availability of data we identified three subpopulations: Asian, OECD, and Latin American subsamples[6]. Tables 5 to 7 in the Appendix 2 describe statistically the variables for regional and economic subsample (see Appendix 1 for code, sample composition, subsample countries list). The main information provided by these Tables was summarized in Figure 1. There it can be seen that main differences in subsamples OECD and Asian reside in KRES and KNRES, while KDUR is higher in the Asian sample. Developed countries (OECD subsample) as well as higher growth rate countries are quite similar each with the full sample averages but not with the Latin American case. In the last one we can observe a high preponderance of KOTHR and lower KDUR as an interesting characteristic jointly with the

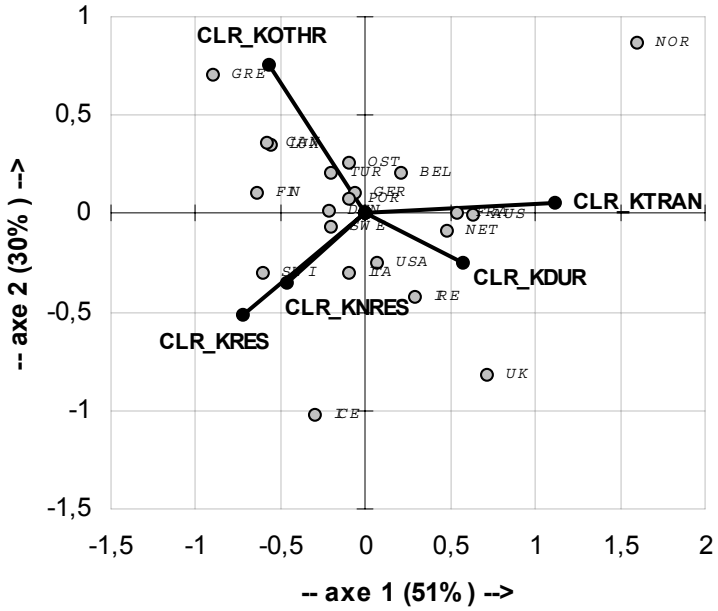
scarce participation of KNRES compared with the full sample averages. Last three columns of Appendix 1 indicate with 1 the inclusion of each observation in subsamples. Beginning with the Asian subsample and after applying the *clr* transformation we started the PCA. This subsample is the smallest with 8 observations and could be a little daring to apply this technique, but this will be done for illustrative purposes. The biplot can be seen in Figure 5. It can be seen that the relationship structure has changed compared with the full sample. Now we have KDUR and KOTHR and KTRAN and KNRES, *vis a vis*, in almost perfectly, negative correlation. KRES is the only that has little correlation with all other components. It seems that when allocating physical capital in sample Asian countries there was a kind of struggle between these components. Higher participation of KDUR was made at expenses of KOTHR, and the same could be said about KTRAN and KNRES. Here amalgamation procedures seem more difficult to justify. In any case, correlation is lower than in full sample case.

**Figure 5. Biplot on the first two principal components (86%) – Asian sample**




In the case of OECD subsample (Figure 6) we obtained similar results as shown in the full sample with, in this case, almost perfectly coincidence between KNRES and KRES.

**Figure 6. Biplot on the first two principal components (81%) – OECD sample**




In Latin American subsample variables' behavior, again, is broadly similar to the full sample analysis. In Figure 9 estimation depicts a higher correlation between KDUR and KTRAN than in other subsamples. However is interesting to compare the possible linear relationship between components across subsamples. Comparing the subsamples compositional rects in the simplexes could better present these differences. This is published in Figure 8a to 8c. We can see that OECD and Latin American subsamples hold a similar compositional rect with a little bias between the same components. Asian subsample, in the other hand, has a different compositional rect tracing a relationship between KEQUIP and KBUILD. However, as observed in Figure 5, we should not amalgamate primitive variables as we did in the other two subsamples because statistical relationship was not preserved. For the sake of exposition, we amalgamated using Figure 5 information KBUILD = KNRES + KOTHR, and we kept KEQUIP as the usual definition. KRES behaves with lower correlation so we separated it and them we estimate a compositional rect on the simplex, as pictured in Figure 8d. Although this information could not be compared directly with the formers it represents more precisely the statistical information obtained by PCA.

**Figure 7. Biplot on the first two principal components (81%) – Latin American sample**

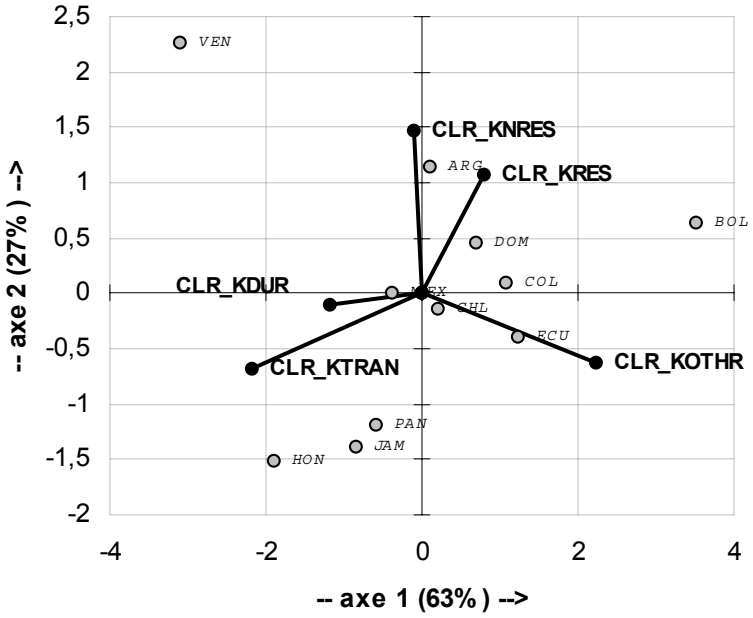
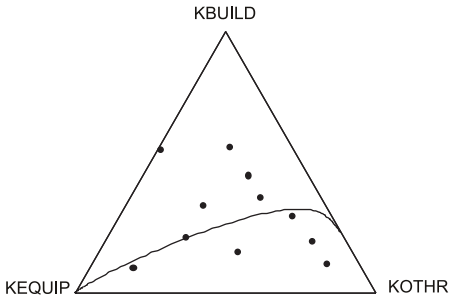
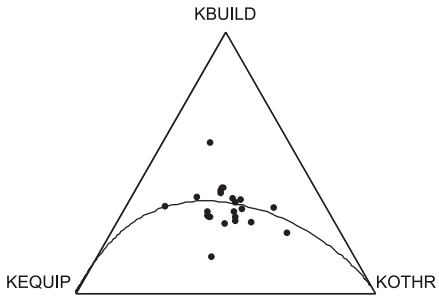



**Figure 8. Compositional Rects for Subsamples**

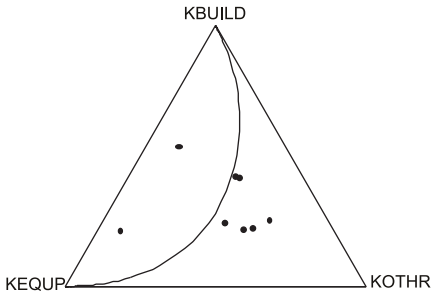
**(a) Latin American Subsample**



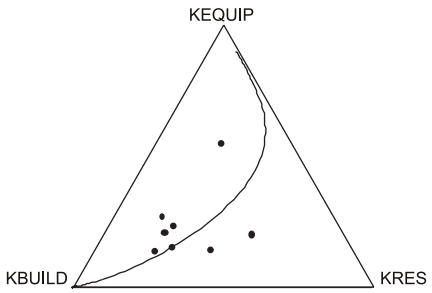

**(b) OECD Subsample**




**(c) Asian Subsample**




**(d) Asian Subsample**



## 4. Preliminary Conclusions and Discussion

We analyzed a static sample of capital per worker composition for understanding the internal changes these compositions have taken place. We distinguished main patterns of behavior as follows:

1) Components of capital allocated in building industries behave differently from those allocated in others activities (especially considering the defined equipment amalgamation).

2) This behavior seems to be different whether we considered geographic and economic sub samples, especially related to one subsample.

3) Sub sample conclusions could be constrained by small sample bias (especially present in the odd sub sample).

4) We conjecture two possible explanations for the observed behavior:

a) Displacement among sectors (presence of collinearity) could be interpreted as a kind of sector struggle for capital allocation. Assigning capital to one sector necessarily implies diminishing capital to another. This report helps to see the direction and affected sectors of these changes.

b) Coincident vertices show sectors that show a joint behavior: they both raise and fall together during the economic process. The observed case of KBUILD (KRES+KNRES) could be better understood as the behavior of two complementary sectors: increment in non-residential construction is made jointly with an increment in the residential counterpart (the dam and the required workers' houses initially exemplified). This is not that clear in the equipment sector. Transportation and durable goods changes show lower correlation between them (different direction and a lesser ray length)[7].

5) In the specific case of Asian subsample we found a dissimilar pattern. We found crossed negative correlation between two sectors, each belonging to a different specific functionality. In this case we cannot rely on the complementarities approach because there was no coincident rays in the PCA. It seems like each sector behaves in a sort of lower inter-sector dependent behavior.

We could mention as future paths of research two main approaches: First, there's no dynamical analysis in this report. It would be interesting to consider how these patterns have changed over the sample period. This could bring some evidence on potential structural breaks or sudden changes in the capital composition over the sample. Second, and especially related with this last proposition, it could be highly motivating the study on how capital composition has influenced the economic growth process. For this purpose, it would

be interesting to test this relationship using the currently available and extensive growth empiric datasets and research papers.

## End Notes

[1] A formalized and stylized framework of these results can be seen in Barceló-Vidal *et al.* (2001).

[2] For an extensive application of additive logistic transformation in Biology see Billheimer *et al.* (1998).

[3] We could rely altogether on the central limit theorem for larger samples too.

[4] See Aitchinson and Greenacre (2001) for an excellent presentation of the issues described in this section.

[5] We could use clustering techniques for identifying statistical subpopulations but for the sake of clarity we considered this more intuitive classification (for a technical analysis of the former perspective see Martín-Fernandez *et al.* (1998)).

[6] As controversial issues could be mentioned that Asian sample includes Japan and OECD sample includes European Economic Community (including Turkey), USA, Canada, and Australia.

[7] Recall Canning and Bennathan (2001) observations on the externality approach to infrastructure research.

## References


Aitchinson, J. (1986), THE STATISTICAL ANALYSIS OF COMPOSITIONAL DATA, Chapman and Hall Ltd., London, 416 p.

Aitchinson, J. and M. Greenacre (2001), "Biplots of Compositional Data", *Working Paper* No. **557**, Departament of Economics and Management, Universitat Pompeu Fabra.

Barceló-Vidal, C., J.A. Martín-Fernández, and V. Pawlowsky-Glahn (2001), "Mathematical Foundations of Compositional Data Analysis", 2001 Annual Conference of the International Association for Mathematical Geology Papers, Kansas Geological Survey, University of Kansas.

Billheimer, D., P. Guttorp , and W.F. Fagan (1998), "Statistical Analysis and Interpretation of Discrete Compositional Data", *NCRSE Technical Report Series* No. **011**.

Canning, D. (2000), "The Contribution of Infrastructure to Aggregate Output", Working Paper Series **2246**, World Bank, Washington.

Canning, D. and E. Bennathan (2000), "The Social Rate of Return of Infrastructure Investments", Policy Research Working Papers **2390**, World Bank, Washington.

De Long, B. and L. Summers (1990), "Equipment Investment and Economic Growth", Working Paper **3515**, National Bureau of Economic Research.

Fry, J.M., T.R.L. Fry, and K.R. McLaren (1996), "Compositional Data Analysis and Zeros in Micro Data", *Applied Economics*, Vol. **32**, 2000, pp. 953-959.

Glaeser, E.L. and J. Gyourko (2001), "Urban Decline and Durable Housing", *Working Paper* **8598**, National Bureau of Economic Research, MA.

Guasch, J.L. and J. Kogan (2001), "Inventories in Developing Countries: Levels and Determinants, a Red Flag on Competitiveness and Growth", *Working Papers Series* **2552**, World Bank, Washington.

Hall, R.E. and C. Jones (1998), "Why Do Some Countries Produce So Much More Output per Worker Than Others?", *Working Paper* **6564**, National Bureau of Economic Research, Cambrige, MA.

Ingram, G. and Z. Liu (1997), "Motorization and Road Provision in Countries and Cities", *Working Paper Series* **1842**, World Bank, Washington.

_____ (1999), "Determinants of Motorization and Road Provision", *Policy Research Working Paper* **2042**, World Bank, Washington.

Jones, C. (1994), "Economic growth and the relative price of capital", *Journal of Monetary Economics* **34**, No. 3, December, pp. 359-382.

Jovanic, B. and R. Rob (1997), "Solow vs. Solow: Machine Prices and Development", *Working Paper* **5871**, National Bureau of Economic Research, Cambrige, MA.

Klingebiel, D. and J. Ruster (2000), "Why Infrastructure Facilities Often Fall Short of Their Objectives", *Working Paper* **2358**, World Bank Institute, Washington.

Martín-Fernández, J.A., C. Barceló Vidal and V. Pawlowsky-Glahn (1998), "A Crtitical Approach to Non-Parametric Classification of Compositional Data", Proceedings of the 5th Conference of the International Federation of Classification Societies, Universitá La Sapienza, Rome.

_____(2000), "Zero Replacement in Compositional Data Sets", Proceedings of the 7th Conference of the International Federation of Classification Societies, Namur, Belgium.

Pearson, K. (1897), "Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs", *Proceedings of the Royal Society* **60**, 489-498.

Randolph, S., Z. Bogetic, and D. Heffley (1996), "Determinants of Public Expenditure on Infrastructure: Transportation and Communication", *Working Paper Series* **1661**, World Bank, Washington.

Reinikka, R. and J. Svensson (1999), "How Inadequate Provision of Public Infrastructure and Services Affects Private Investment", *Policy Research Working Paper* **2262**, The World Bank, Washington.

Temple, J, and H.S. Voth (1998), "Human capital, equipment investment, and industrialization", *European Economic Review*, July, 42(7), 1343-1362.

## Appendix 1. Raw Data, Country List, and Subsample selection

| Country | KTRAN | KOTHR | KDUR | KRES | KNRES | Asian | OECD | LA |
|---|---|---|---|---|---|---|---|---|
| ARG | 0,01246135 | 0,18419813 | 0,08330162 | 0,28025142 | 0,43978747 | | | 1 |
| AUS | 0,05539246 | 0,17318537 | 0,21913061 | 0,2837683 | 0,26852325 | | | |
| OST | 0,02477246 | 0,24183522 | 0,20386712 | 0,26276062 | 0,26676458 | | 1 | |
| BEL | 0,03546181 | 0,22636425 | 0,22398637 | 0,29206138 | 0,22212619 | | 1 | |
| BOL | 0,00033137 | 0,76180208 | 0,06290189 | 0,11253535 | 0,0624293 | | | 1 |
| BOT | 0,01612742 | 0,12913198 | 0,27589736 | 0,26785248 | 0,31099076 | | | |
| CAN | 0,0201624 | 0,29137834 | 0,09466554 | 0,39243929 | 0,20135444 | | 1 | |
| CHL | 0,02636323 | 0,38061589 | 0,07623877 | 0,35716859 | 0,15961353 | | | 1 |
| COL | 0,00978964 | 0,51005124 | 0,05758199 | 0,26849491 | 0,15408222 | | | 1 |
| DEN | 0,02396707 | 0,20292309 | 0,16510387 | 0,33819544 | 0,26981053 | | 1 | |
| DOM | 0,00967782 | 0,29362318 | 0,08772119 | 0,47472714 | 0,13425066 | | | 1 |
| ECU | 0,00965209 | 0,6418866 | 0,05243228 | 0,19443052 | 0,10159851 | | | 1 |
| FIN | 0,01395756 | 0,22718412 | 0,17305838 | 0,30226796 | 0,28353198 | | 1 | |
| FRA | 0,05011375 | 0,17899396 | 0,22365749 | 0,29274802 | 0,25448677 | | 1 | |
| GER | 0,02784907 | 0,21710483 | 0,18154098 | 0,31971812 | 0,253787 | | 1 | |
| GRE | 0,01224733 | 0,39337231 | 0,12214802 | 0,30961552 | 0,16261682 | | 1 | |
| HON | 0,17441017 | 0,19448265 | 0,41739048 | 0,12235131 | 0,09136539 | | | 1 |
| HKG | 0,13460477 | 0,05222804 | 0,41383056 | 0,21666399 | 0,18267264 | 1 | | |
| ICE | 0,01838309 | 0,07113572 | 0,1124462 | 0,61293213 | 0,18510285 | | 1 | |
| IND | 0,01547601 | 0,37203898 | 0,135968 | 0,25102974 | 0,22548727 | 1 | | |
| IRE | 0,03364768 | 0,12153323 | 0,22062188 | 0,32058329 | 0,30361391 | | 1 | |
| ISR | 0,00980037 | 0,05393613 | 0,19159013 | 0,49130609 | 0,25336729 | 1 | | |
| ITA | 0,02644213 | 0,15211524 | 0,1668867 | 0,45821458 | 0,19634135 | | 1 | |
| IVC | 0,01872635 | 0,23066937 | 0,18280751 | 0,38428975 | 0,18350701 | | | |
| JAM | 0,07162015 | 0,28708863 | 0,26399592 | 0,33337419 | 0,04392112 | | | 1 |
| JAP | 0,04640137 | 0,33477738 | 0,19002559 | 0,21482513 | 0,21397052 | 1 | | |
| KOR | 0,01779361 | 0,20579034 | 0,12017969 | 0,20110999 | 0,45512637 | 1 | | |
| LUX | 0,0156355 | 0,27766447 | 0,17902026 | 0,27962057 | 0,2480592 | | 1 | |
| MEX | 0,03153946 | 0,24750216 | 0,19644783 | 0,34834809 | 0,17616246 | | | 1 |
| NET | 0,04594907 | 0,16586137 | 0,22047257 | 0,29830014 | 0,26941685 | | 1 | |
| NEW | 0,0415393 | 0,48619287 | 0,2041567 | 0,18753183 | 0,0805793 | | | |
| NOR | 0,14501814 | 0,2836095 | 0,25055373 | 0,15073494 | 0,17008368 | | 1 | |
| PAN | 0,07867878 | 0,49085157 | 0,1563809 | 0,1131978 | 0,16089096 | | | 1 |
| POR | 0,02887859 | 0,20886015 | 0,14142183 | 0,51749476 | 0,10334467 | | 1 | |
| SLE | 0,06284731 | 0,40218092 | 0,2625121 | 0,11090145 | 0,16155822 | | | |
| SWE | 0,02467155 | 0,1913277 | 0,15898652 | 0,37037939 | 0,25463484 | | 1 | |
| SWI | 0,01336942 | 0,15277321 | 0,16963094 | 0,33190497 | 0,33232146 | | 1 | |
| SYR | 0,03503957 | 0,19707791 | 0,10732429 | 0,38630771 | 0,27425052 | 1 | | |
| TAI | 0,01932647 | 0,27285552 | 0,24879952 | 0,16109129 | 0,2979272 | 1 | | |
| THA | 0,01349893 | 0,35232416 | 0,19618864 | 0,19951711 | 0,23847117 | 1 | | |
| TUR | 0,0220548 | 0,232157 | 0,19650091 | 0,26115564 | 0,28813165 | | 1 | |
| UK | 0,04156811 | 0,07374036 | 0,29897192 | 0,32416378 | 0,26155583 | | 1 | |
| USA | 0,03288553 | 0,15702049 | 0,16461581 | 0,42187362 | 0,22360455 | | 1 | |
| VEN | 0,03544679 | 0,00691426 | 0,18706483 | 0,27838119 | 0,49219292 | | | 1 |

## Appendix 2. Descriptive Statistics

### Table 4. Descriptive statistics of variables (Full sample)

| | KTRAN | KOTHR | KDUR | KRES | KNRES |
|---|---|---|---|---|---|
| Mean | 0,036445 | 0,25746273 | 0,18313694 | 0,29765044 | 0,22530489 |
| Geometric Mean | 0,02506905 | 0,20830026 | 0,16591767 | 0,27596401 | 0,20329306 |
| Median | 0,02556784 | 0,22677419 | 0,18217425 | 0,2924047 | 0,22454591 |
| Standard Dev. | 0,03599049 | 0,15205473 | 0,07943213 | 0,11170121 | 0,09564047 |
| Sample Variance | 0,00129532 | 0,02312064 | 0,00630946 | 0,01247716 | 0,0091471 |
| Kurtosis | 6,43488218 | 2,21173911 | 1,78363416 | 0,49024098 | 1,08410137 |
| Asymmetry Coefficient | 2,48136527 | 1,24220285 | 0,88428386 | 0,49079849 | 0,63036541 |
| Rank | 0,17407879 | 0,75488782 | 0,3649582 | 0,50203068 | 0,4482718 |
| Minimum | 0,00033137 | 0,00691426 | 0,05243228 | 0,11090145 | 0,04392112 |
| Maximum | 0,17441017 | 0,76180208 | 0,41739048 | 0,61293213 | 0,49219292 |
| Sample | 44 | 44 | 44 | 44 | 44 |

### Table 5. Descriptive statistics of variables (Asian sample)

| Measure | KTRAN | KOTHR | KDUR | KRES | KNRES |
|---|---|---|---|---|---|
| Mean | 0,03649264 | 0,23012856 | 0,2004883 | 0,26523138 | 0,26765912 |
| Geometric Mean | 0,02503058 | 0,18493039 | 0,1835744 | 0,24805389 | 0,25827995 |
| Median | 0,01856004 | 0,23932293 | 0,19080786 | 0,21574456 | 0,24591923 |
| Standard Dev. | 0,04149613 | 0,12661774 | 0,09808735 | 0,11348102 | 0,08370305 |
| Sample Variance | 0,00172193 | 0,01603205 | 0,00962113 | 0,01287794 | 0,0070062 |
| Kurtosis | 5,9534168 | -1,27345717 | 3,29059339 | 1,28630302 | 4,18253738 |
| Asymmetry Coefficient | 2,38141769 | -0,49205989 | 1,66029941 | 1,4867932 | 1,86597524 |
| Rank | 0,1248044 | 0,31981093 | 0,30650627 | 0,3302148 | 0,27245373 |
| Minimum | 0,00980037 | 0,05222804 | 0,10732429 | 0,16109129 | 0,18267264 |
| Maximum | 0,13460477 | 0,37203898 | 0,41383056 | 0,49130609 | 0,45512637 |
| Sample | 8 | 8 | 8 | 8 | 8 |

### Table 6. Descriptive statistics of variables (OECD sample)

| Measure | KTRAN | KOTHR | KDUR | KRES | KNRES |
|---|---|---|---|---|---|
| Mean | 0,03392512 | 0,20191143 | 0,18510894 | 0,3400444 | 0,23901011 |
| Geometric Mean | 0,02814859 | 0,18763072 | 0,17894046 | 0,32729354 | 0,23205531 |
| Median | 0,02644213 | 0,20292309 | 0,17902026 | 0,31971812 | 0,25448677 |
| Standard Dev. | 0,02815633 | 0,07441855 | 0,04795718 | 0,09871245 | 0,05373148 |
| Sample Variance | 0,00079278 | 0,00553812 | 0,00229989 | 0,00974415 | 0,00288707 |
| Kurtosis | 12,9807359 | 1,14298985 | 0,45863447 | 2,31307415 | 0,67102688 |
| Asymmetry Coefficient | 3,3251643 | 0,47771509 | 0,25371217 | 1,13480311 | -0,76599291 |
| Rank | 0,13277081 | 0,32223659 | 0,20430638 | 0,46219719 | 0,22897679 |
| Minimum | 0,01224733 | 0,07113572 | 0,09466554 | 0,15073494 | 0,10334467 |
| Maximum | 0,14501814 | 0,39337231 | 0,29897192 | 0,61293213 | 0,33232146 |
| Sample | 21 | 21 | 21 | 21 | 21 |

**Table 7. Descriptive statistics of variables (Latin American sample)**

| Measure | KTRAN | KOTHR | KDUR | KRES | KNRES |
|---|---|---|---|---|---|
| Mean | 0,04181553 | 0,36354695 | 0,14922343 | 0,26211459 | 0,1832995 |
| Geometric Mean | 0,01889154 | 0,25022558 | 0,11887624 | 0,23574594 | 0,14310769 |
| Median | 0,02636323 | 0,29362318 | 0,08772119 | 0,27838119 | 0,15408222 |
| Standard Dev. | 0,0508813 | 0,22049459 | 0,11290381 | 0,11664169 | 0,14653929 |
| Sample Variance | 0,00258891 | 0,04861786 | 0,01274727 | 0,01360528 | 0,02147376 |
| Kurtosis | 4,55086229 | -0,27698416 | 2,18258286 | -0,59986802 | 1,45741076 |
| Asymmetry Coefficient | 2,04750349 | 0,36063323 | 1,50560361 | 0,13818198 | 1,56849736 |
| Rank | 0,17407879 | 0,75488782 | 0,3649582 | 0,36219179 | 0,4482718 |
| Minimum | 0,00033137 | 0,00691426 | 0,05243228 | 0,11253535 | 0,04392112 |
| Maximum | 0,17441017 | 0,76180208 | 0,41739048 | 0,47472714 | 0,49219292 |
| Sample | 11 | 11 | 11 | 11 | 11 |

## Appendix 2. PCA estimations

**Table 8. Eigenvalues and eigenvectors (based on the covariance matrix)**

| Eigenvalues | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Value | 1,0127 | 0,5676 | 0,1585 | 0,1144 | 0,0000 |
| % of variability | 0,5465 | 0,3063 | 0,0855 | 0,0617 | 0,0000 |
| Cumulative % | 0,5465 | 0,8528 | 0,9383 | 1,0000 | 1,0000 |
| **Vectors** | **1** | **2** | **3** | **4** | **5** |
| CLR_ KTRAN | -0,6548 | -0,4885 | 0,1557 | -0,3293 | 0,4472 |
| CLR_ KOTHR | 0,7235 | -0,5115 | -0,0649 | -0,1032 | 0,4472 |
| CLR_ KDUR | -0,1857 | 0,0070 | -0,3316 | 0,8096 | 0,4472 |
| CLR_ KRES | 0,1153 | 0,4388 | 0,7656 | 0,0892 | 0,4472 |
| CLR_ KNRES | 0,0017 | 0,5542 | -0,5248 | -0,4663 | 0,4472 |

*Correlations between initial variables and principal factors*

| | factor 1 | factor 2 | factor 3 | factor 4 | factor 5 |
|---|---|---|---|---|---|
| CLR_ KTRAN | -0,8609 | -0,4808 | 0,0810 | -0,1455 | 0,0000 |
| CLR_ KOTHR | 0,8826 | -0,4671 | -0,0313 | -0,0423 | 0,0000 |
| CLR_ KDUR | -0,5237 | 0,0148 | -0,3699 | 0,7673 | 0,0000 |
| CLR_ KRES | 0,2493 | 0,7104 | 0,6550 | 0,0648 | 0,0000 |
| CLR_ KNRES | 0,0035 | 0,8472 | -0,4240 | -0,3200 | 0,0000 |