

Avanzini, Diego B.

Working Paper

Designing composite entrepreneurship indicators: An application using consensus PCA

WIDER Research Paper, No. 2009/41

Provided in Cooperation with:

United Nations University (UNU), World Institute for Development Economics Research (WIDER)

Suggested Citation: Avanzini, Diego B. (2009) : Designing composite entrepreneurship indicators: An application using consensus PCA, WIDER Research Paper, No. 2009/41, ISBN 978-92-9230-212-2, The United Nations University World Institute for Development Economics Research (UNU-WIDER), Helsinki

This Version is available at:

<https://hdl.handle.net/10419/45162>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



UNITED NATIONS
UNIVERSITY

UNU-WIDER

World Institute for Development
Economics Research

Research Paper No. 2009/41

Designing Composite Entrepreneurship Indicators

An Application Using Consensus PCA

Diego B. Avanzini*

June 2009

Abstract

Existing indicators of entrepreneurial activity (such as the Global Entrepreneurship Monitor, Entrepreneurship Barometer, FORA's Entrepreneurship Index, OECD and Economic Commission's sets of indicators, among others) and several variables that have been considered good proxies for entrepreneurship during the last decades seem to be unsuitable to capture the complex relationship among economic, social, and demographic factors driving entrepreneurial activity.

In order to suggest a consistent methodology for measuring entrepreneurship, we review some of the most well-known theoretical dimensions of entrepreneurship and a selection of associated indicators is proposed. Indicators and measures are grouped under theoretical categories and a set of entrepreneurship indicators is constructed using multivariate statistical analysis (Consensus PCA based on NIPALS, with an extension of Probability PCA for dealing with missing values) for a panel of developed and developing countries.

Keywords: entrepreneurship, index, principal components analysis

JEL classification: L26, M13, C82

Copyright © UNU-WIDER 2009

* Catholic University of Chile, Santiago, email: Diego.AVANZINI@cepal.org

This study has been prepared within the UNU-WIDER project on Entrepreneurship and Development (Promoting Entrepreneurial Capacity), directed by Wim Naudé.

UNU-WIDER gratefully acknowledges the financial contributions to the project by the Finnish Ministry for Foreign Affairs, and the financial contributions to the research programme by the governments of Denmark (Royal Ministry of Foreign Affairs), Finland (Finnish Ministry for Foreign Affairs), Norway (Royal Ministry of Foreign Affairs), Sweden (Swedish International Development Cooperation Agency—Sida) and the United Kingdom (Department for International Development).

ISSN 1810-2611

ISBN 978-92-9230-212-2

The World Institute for Development Economics Research (WIDER) was established by the United Nations University (UNU) as its first research and training centre and started work in Helsinki, Finland in 1985. The Institute undertakes applied research and policy analysis on structural changes affecting the developing and transitional economies, provides a forum for the advocacy of policies leading to robust, equitable and environmentally sustainable growth, and promotes capacity strengthening and training in the field of economic and social policy making. Work is carried out by staff researchers and visiting scholars in Helsinki and through networks of collaborating scholars and institutions around the world.

www.wider.unu.edu

publications@wider.unu.edu

UNU World Institute for Development Economics Research (UNU-WIDER)
Katajanokanlaituri 6 B, 00160 Helsinki, Finland

Typescript prepared by the author

The views expressed in this publication are those of the author(s). Publication does not imply endorsement by the Institute or the United Nations University, nor by the programme/project sponsors, of any of the views expressed.

Designing Composite Entrepreneurship Indicators: An Application using Consensus PCA*

Diego B. Avanzini[†]
Catholic University of Chile

December 7, 2008

Abstract

Existing indicators of entrepreneurial activity (such as Global Entrepreneurship Monitor, Entrepreneurship Barometer, FORA's Entrepreneurship Index, OECD and Economic Commission's sets of indicators, among others) and several variables that have been considered good proxies for entrepreneurship during last decades seem to be not suitable to capture the complex relationship among economic, social, and demographic factors driving entrepreneurial activity.

In order to suggest a consistent methodology for measuring entrepreneurship, we review some of the most well-known theoretical dimensions of entrepreneurship and a selection of associated indicators is proposed. Indicators and measures are grouped under theoretical categories and a set of entrepreneurship indicators are constructed using multivariate statistical analysis (Consensus PCA based on NIPALS, with an extension of Probability PCA for dealing with missing values) for a panel of developed and developing countries.

1 Introduction

For more than two hundred and fifty years we have been using the expression "Entrepreneurship" but without a single, unambiguous meaning that adequately describes its nature. Furthermore, ideas associated with entrepreneurship are continuously evolving, so it depends on the historical context we are analyzing it. And if we do not agree on what it is, we could difficultly achieve a consensus on how to measure it. In 1933, J. A. Schumpeter wrote:

"... as long as we are unable to put our arguments into figures, the voice of our science, although occasionally it may help to dispel gross errors, will never be heard by practical men." (Schumpeter, 1933)

Schumpeter stated the general problem of measuring, here we bring it to the field of entrepreneurial activity. Entrepreneurship has been understood as an "invisible force" leading economic efforts, generating knowledge, innovating, attracting employees, improving economic growth, and so on. These ideas come together with several definitions and concepts that highlight different dimensions of entrepreneurship. The problem is that none of them fully achieve the aim to define entrepreneurship. Such a rich expression poses also a major problem from the measurement point of view, preventing the adequate communication of this important phenomenon. "Practical men" like to have numbers to show and to think about. Without measuring entrepreneurial activity, we will hardly communicate the relevance of policy-making on an entrepreneurship-oriented base to policy- and decision-makers.

*This paper was presented at **WIDER Project Workshop on Entrepreneurship and Economic Development: Concepts, Measurements, and Impacts**, Helsinki, 21 - 23 August 2008. I am very grateful to Workshop's assistants and Wim Naude for their comments, and also to Carolina Serpell for reviewing the document's wording. As usual, remaining errors are the author liability.

[†]PhD. in Economics student, field of specialization Quantitative and Applied Economics. E-mail: dbavanzi@uc.cl, diego_avanzini@yahoo.com

Literature on entrepreneurship comes from different areas, and each author highlights different features of the phenomenon. Our proposal consists of capturing the main and widely-accepted common features and constructing measures of entrepreneurship based on robust statistical methods. The final outcome of our work is the Composite Entrepreneurship Indicator (CEI) that is obtained applying a set of theoretical and empirical rules developed in Avanzini (2008), and its methodological framework. We also discuss the use of Consensus Principal Component Analysis (CPCA) in this context. One additional problem relates to data scarcity. This problem is particularly hard to overcome in developing countries. In what follows, we suggest the use of an extension of Probability Principal Component Analysis (PPCA) by Roweis (1998) to deal with missing values, and we also discuss other alternatives to strengthen the dataset quality (Section 4).

Beyond the introduction (Section 1) and the concluding remarks (Section 8), the document is organized according to the six stages detailed in Avanzini (2008), namely setting a definition of entrepreneurship (Section 2), selecting indicators (Section 3), arranging data (Section 4), designing the composite indicator (Section 5), estimating the set of indicators (Section 6), and discussing the outcomes (Section 7). In those sections, we review the relevant literature on entrepreneurship and tackle the problem of its measurement. We give the theoretical background of the statistical tools in the Appendix, together with a description of the database we use in estimations.

2 Defining Entrepreneurship

"Scholars have dedicated almost three centuries to the attempt to define the concept of entrepreneurship. The lack of consensus may, in part, be due to the fact that entrepreneurship is not neatly contained within any single academic domain. Indeed, many disciplines have contributed their perspectives on the concept of entrepreneurship, including Psychology (Shaver and Scott, 1991), Sociology (Reynolds, 1991; Thorton, 1999), Economics (Cantillon, 1755; Marshall, 1890; Knight, 1921; Schumpeter, 1934, 1949) and Management (Stevenson, 1985). Given the heightened interest in entrepreneurship in recent years, it is unlikely this multi-disciplinary interest will diminish any time soon." (OECD, 2006).

This paragraph gives us a clear picture of the problem we are facing: *Entrepreneurship is a multifaceted and heterogeneous activity* (Audretsch and Thurik, 2001; Audretsch, 2002).

The French economist Richard Cantillon is generally accredited as being the first to coin the term "entrepreneurship" circa 1730¹. Loosely, he defined entrepreneurship as *self-employment* of any sort involved in a process of *bearing the risk to organize* factors of production to deliver a product or service demanded by the market. Knight (1921) also stressed the *risk-bearing* dimension of entrepreneurship. Alfred Marshall (1890) identified entrepreneurship as a crucial factor of production alongside land, capital and labour. Other authors have stressed the importance of the entrepreneur in different roles: manager (Say, 1803), speculator (Von Mises, 1949), coordinator and arbitrageur (Walras, 1954). Kirzner (1973) stressed the *profit-seeking* activity through *market arbitrage*, while Penrose (1959) highlights the *opportunity identification* activity. These ideas can be grouped under the "*old school*" with the following features: risk-bearing, organization / management, arbitrage / coordination / speculation, profit-seeking, and self-employment.

Joseph Schumpeter (1934) opened a new dimension of what we consider entrepreneurship: *innovation*. Entrepreneurship is thus seen as the process of identifying, developing, and bringing forward new innovative ways of doing things for the exploitation of commercial opportunities. And this is the definition that gains major acceptance in last decades. Since Schumpeter, entrepreneurship is equated with the concept of innovation applied to a business context. He defined entrepreneurs as *innovators* who implement entrepreneurial change within markets, where entrepreneurial change has five manifestations: (i) the introduction of a new (or improved) good; (ii) the introduction of a new method of production; (iii) the opening of a new market; (iv) the conquest of a new source of supply of new materials or parts; and (v) the re-engineering or re-organization of business management processes, or the carrying out of the new organization of any industry. Drucker (1985) built his idea of entrepreneurship on Schumpeter's entrepreneurial innovation process posing that entrepreneurship is the act of innovation involving endowing existing resources with

¹The word *entrepreneur* itself derives from the French verb "*entreprendre*", meaning "to undertake". There is a general belief that the first drafts of R. Cantillon date from 1730 approximately, but the first edition of his manuscripts was published in 1755 (Cantillon, 1755).

new wealth-producing capacity. Although Schumpeter's view of entrepreneurship is a tempting one and is widely accepted nowadays, it still retains some ambiguity that has meant the debate regarding a definition of entrepreneurship continues. Indeed some authors (Drucker, 1985; Lumpkin and Dess, 1996) have argued that entrepreneurship reflects merely the creation of a new organization, or that the essential act of entrepreneurship is new entry. This new entry can be accomplished by entering new or established markets with new or existing goods or services. New entry is the act of launching a new venture, either by a start-up firm, through an existing firm, or via "internal corporate venturing". We may say that these ideas involving innovation and the "old school" characteristics can be put together under a "*second generation*" concept.

In recent years, many national governments and international organizations have been encouraging analysts and researchers to unify criteria and try to set up a general framework to study and evaluate entrepreneurship. This "*third generation*" of studies on entrepreneurship tries to integrate three dimensions: (i) entrepreneurship, as it is defined by the "second generation" literature; (ii) the integration of activities in a process (not just separate activities but complete plans of action aimed to seek certain goals including the organization survival), and (iii) the need for its measurement. For example, Shane and Venkataraman (2000) said that the field of entrepreneurship involves the study of sources of opportunities; the processes of discovery, evaluation, and exploitation of opportunities; and the set of individuals who discover, evaluate, and exploit them. Ireland et al. (2003) posed that entrepreneurship is a context-dependent social process through which individuals and teams create wealth by bringing together unique packages of resources to exploit marketplace opportunities. Finally, and without exhausting the list of examples, the Commission of the European Communities (2003) arrived to the conclusion that entrepreneurship is the mindset and process to create and develop economic activity by blending risk-taking, creativity and / or innovation with sound management, within a new or an existing organization.

Ahmad and Seymour (2008) said that all these kind of concepts and definitions can be analyzed from a philosophical perspective, the so-called "top-down approach". Under this approach, entrepreneurship is studied with little concern for measurement. On the other extreme, the "bottom-up approach" devoted all its effort to equate entrepreneurship to a specific empirical measure or set of measures, avoiding the discussion of entrepreneurship definitions altogether. The problem with this approach is that provided there is no relevant concept or definition that guides the selection of the indicators, the chosen measures are those based on the most readily available statistics, and only rarely do authors attempt to justify or explain how the measures represent entrepreneurship.

The definition of entrepreneurship we will adopt for our composite indicator is the one developed by Ahmad and Seymour (2008). For them the *entrepreneur* is an economic agent that "[...] is simultaneously looking back to the resources (and combining them in new and creative ways) and forward to markets (and perceiving new or unmet opportunities). The entrepreneur perceives and recognizes a fit between the two, a capability and process referred to as innovating. The entrepreneur's activities occur within a business context, which includes industry structures, competition, and national economic structures. This business context is impacted in turn by wider environmental considerations, which include the economic, political, legal, social, cultural, and natural settings. In undertaking such *entrepreneurial activities*, the entrepreneur is endeavouring to create value" (Ahmad and Seymour, 2008, pp. 5). The entrepreneurship definition we adopt is the following:

"Entrepreneurship is the phenomenon associated with the entrepreneurial activity, i.e. the enterprising human action in pursuit of the generation of value, through the creation or expansion of economic activity, by identifying and exploiting new products, processes or markets." (Ahmad and Seymour, 2008).

This definition does not encompass all the possible dimensions that are discussed in the specialized literature, but it constitutes a good approximation in order to guide our procedure. Depending on the definition the analyst adopts the outcomes and the questions she can answer will differ. This definition is very suitable for our application specially because it was developed thinking in the necessity of measuring entrepreneurship, something that would facilitate the selection of sub-indicators in the next stage.

3 Selecting Dimensions and Indicators

Policy-makers are interested in the determinants and impacts of entrepreneurship: they need to know what fosters entrepreneurial activity, and which are the effects and spillovers it causes. We assume policies are driven by certain goals related to entrepreneurship and policy-makers need indicators to inform them how these policies affect entrepreneurship and achieve the goals. Provided the multifaceted nature of entrepreneurship, even in cases where a fairly clear definition has been enunciated, it is difficult to find a measurement tool that matches the terminology that has been chosen². Existing indicators of entrepreneurial activity (such as Global Entrepreneurship Monitor, Entrepreneurship Barometer, FORA's Entrepreneurship Index, OECD and Economic Commission's sets of indicators, among others) and several variables that have been considered good proxies for entrepreneurship during last decades seem to be not so adequate to capture the complex relationship between economic, social, and demographic factors driving entrepreneurial development.

The wide range of concepts involved in the entrepreneurship definition (that in short constitute its determinants and impacts), and the variety of policy goals and the way in which they can be measured force us not to measure the phenomenon by mean of a lonely indicator but rather a set of them. The chosen practical measure of entrepreneurship will ultimately depend on the nature of the policy objective³. Provided there is neither fully objective way of selecting the relevant sub-indicators, nor a certain way to group them if it would be necessary, we start designing a *scoreboard*, an idea that has appealed to many managers and policy-makers in recent years⁴ and that we adopt here⁵. One of the advantages of the scoreboard is that it is a guide to select and collect available data, and a checklist for not yet available data that should be collected. On the other hand, one of its disadvantages is that it has so many indicators that the likelihood of 'garbage in-garbage out' curse is accentuated. Also having many indicators may be as bad as having a few provided measurement errors are more likely, the relationship among them would be less clear, their own variability could lead us to misunderstand the phenomenon variability, etc. Variables and sub-indicators have been selected⁶ on the basis of (i) their analytical soundness, (ii) measurability, (iii) relevance to the phenomenon being measured, and (iv) relationship to each other.

We find helpful to disaggregate the problem in several dimensions following the way literature have treated them⁷, and we propose the following Scoreboard as the basis for measuring entrepreneurship. Its basic structure is presented in Figure 1. This scoreboard captures the main dimensions of the entrepreneurial activity, facilitating issues and determinants, and its manifestations and impacts. We focus our attention in

²For example, the European Commission has defined entrepreneurship as "the mindset and process needed to create and develop economic activity by blending risk-taking, creativity and/or innovation with sound management, within a new or existing organization". While conceptually appealing, it would be difficult to convey this notion on a questionnaire in a way that would invite consistent interpretation by all respondents (Ahmad and Seymour, 2008).

³For example, if policy-makers are interested in employment creation, they may focus on a measure that seems most directly linked to jobs, such as self-employment or new firm creation, no matter what the size or growth rate of the firm. If the policy objective is competitiveness or productivity growth, however, a measure of entrepreneurship that distinguishes high growth or innovative firms may be preferred. In this case, the firm population of interest may exclude zero-employee firms (self-employment), or even very small firms, from the population of young businesses in order to get a better count of the growth business population. Relevant measures will also depend on the national context and the structure of the business population.

⁴See Kaplan and Norton (1992, 1993, 1995, 1996a, 1996b) for managerial applications. Kaplan and Norton are considered the fathers of the so-called Balanced Scorecard, a scoreboard-based methodology created to manage enterprises on a strategy-oriented base. Related to policy-making applications: see Nilsson (1987, 2000, 2003), OECD (1998) and Petit et al. (1996), among others, for composite leading indicators for the OECD members; SEIFA (2001, 2006), for Australian social policies; VaLUENTIS (2006), for UK human development policies; Hung (2003) for Hong Kong leading indicators; OECD/Eurostat (2003, 2007), for regional entrepreneurship policies; Fukuda and Onodera (2001), for composite coincident economic indicators for Japan; and United Nations (2001) for sustainable development indicators, among many others.

⁵Our Scoreboard uses some ideas that can be found in the *entrepreneurship* (see Leitão da Silva Martins, 2007; Behrens, 2007; OECD/Eurostat, 2007; Ahmad and Seymour, 2008; Ahmad and Hoffmann, 2008), *innovation* (see Zabala-Iturriagoitia et al., 2007), and *investment* literature (see Statistics Netherlands, 2007).

⁶See Freudenberg (2003).

⁷The approach is not new: Leitão (2007) proposes a scoreboard for measuring entrepreneurship that defines six categories: enterprises, human resources, innovation, social economy, initiative, and knowledge. These categories have sub-indicators aimed to measure different aspects of each category. Some of them are not directly measurable but instead proxy variables can be used. OECD's Entrepreneurship Indicators Programme (EIP) is another example of multilevel categorization in such a scoreboard layout (see Davis, 2007; Ahmad and Seymour, 2008; Ahmad and Hoffmann, 2008). The EIP has settled out a scoreboard (the "OECD-Eurostat EIP Indicator Framework, 2007") that divides the variables and indicators under three aggregates, namely determinants, entrepreneurial performance, and impact. Each of these aggregates are divided in a multi-level multi-block structure. For example, the scoreboard proposes to order determinants under six categories: regulatory framework, market conditions, access to finance, R&D and technology, entrepreneurial capabilities, and culture.

seven dimensions:

1. *Entrepreneurial Activity*: this category includes three dimensions, namely (i) Firm Dynamics, (ii) Firm Survival, and (iii) Ownership. The indicators included under Firm Dynamics and Firm Survival are the mostly recognized entrepreneurial activity proxies. Ownership is included to avoid confusion between entrepreneurship and management: both concepts are very closely related, given entrepreneurship involves management, but a key difference between them is the ownership (that can be translated into a *risk-bearing* activity).
2. *Employment*: describes the impact of entrepreneurship on employment, measured by the number of employees associated with new enterprises creation and exit. Employment is one of the universally claimed beneficial effects of entrepreneurship and should be an important manifestation of its existence.
3. *Economic Activity*: beyond employment, other areas of the economic activity may be reflecting the entrepreneurship development: increasing sales amount, small and medium enterprises creation, international trade, number and capitalization of enterprises in the stock market, etc.
4. *Entrepreneurship Spirit, Culture, and Initiative*: people engaged in entrepreneurial activities have some particular characteristics that make them unique, e.g. entrepreneurial potential and propensity, particular skills, reasons to becoming an entrepreneur, and contact with other entrepreneurs.
5. *Barriers to Entrepreneurial Activity and Business Environment*: entrepreneurship arises in a certain time-location coordinate and by its nature involves the integration of many aspects that lead it to a successful performance. The pro-entrepreneurship characteristics of the business environment, the available resources (financial, physical, human, etc.), and the institutions supporting entrepreneurship play a decisive role in entrepreneurial success.
6. *Knowledge Procurement*: following the Schumpeterian association between entrepreneurship and knowledge generation, we attempt to include here those activities oriented to support the knowledge procurement system (human resources, investment in R&D, R&D activities, etc.).
7. *Innovation*: the final outcome of the knowledge procurement system, that may show up in a variety of ways: new products, new markets, new processes, new uses of existing products, etc. This is the way Schumpeterian's entrepreneurship should manifest.

These categories are not exhaustive, and adopting them is in no way a restricting characteristic of our approach provided the techniques proposed in this paper are intended to allow modifications, improvements, and widening, to set up better measures. Nonetheless we have to get a reasonably-sounded and reliable dataset to estimate the CEI and this is our proposal⁸.

4 Data Selection and Description

First of all, we need to associate each dimension selected in the previous stage with a set of variables reflecting its nature. Then we need to add variables to characterize (pros and cons) and influence the particular dimension. Data selection is a hard job: data is wide-spread (websites, books, yearbooks, outlooks, papers, and other documents), and its quality and reliability are difficultly determined. In most of the cases, there exists several well-known data repositories⁹ that are the usual data sources in entrepreneurship studies.

⁸As was previously discussed, there are many ways for characterizing and grouping characteristics and dimensions of entrepreneurship; some of them, due to lack of relevant data, have been avoided, e.g. social economy activities, entrepreneurial talent development and mobility, and the disaggregation of the referenced categories in more specific ones.

⁹Perhaps the most well-known is the GEM (*Global Entrepreneurship Monitor*) Project (since 1998). In Europe there are several information sources: Eurostat's "*Factors of Business Success*" (FoBS) survey; the European Commission's *Eurobarometer* (Gallup Europe, 2002), OECD/Eurostat Entrepreneurship Indicators Programme (2007-2008). There are also some governmental projects: Canada's periodic survey of small and medium enterprises (SME) access to finance; the United States Federal Reserve's periodic survey of small business finances (SSBF); the University of Warwick's first major study of SME finances in the United Kingdom; FORA (the Centre for Business and Economic Research under the Danish Ministry for Economic and Business Affairs); Statistics Netherlands (2007). Finally, data repositories such as The World Bank's World Development

While few, if any, meet all the requirements of analysts and policymakers for internationally-comparable data, there are numerous statistics relating to entrepreneurship already produced by governmental, quasi-governmental and private institutions. Many of these data sets are purely national and some focus only on special niche activities or a specific subset of the population¹⁰.

We have selected a set of more than 70 variables coming from several sources, for a sample of 69 countries (see Table 1 for a list of countries, and Tables 3 and 7 for a list of variables). The list of variables and its description, separated by source, can be found in Appendix A. These variables present some problems such as periodicity, availability, "outlying" observations, clumping, truncation, and relationship among them, that are discussed in what follows.

4.1 Dealing with Data Scarcity

Data availability is an important issue in the design of any indicator. Usually, developing countries are less represented in datasets, biasing the estimation results towards developed countries. Also there is a problem with the periodicity: not all variables are collected on the same time base, or they have time gaps, if not for all countries, at least for some of them. Basically, the datasets are unbalanced (different number of observations for each country) but the statistical methods we will implement need balanced datasets.

From a practical perspective, we use a combination of two approaches to deal with data scarcity: the *first approach* consists of averaging variables for each country over two periods, namely 1998-2001 and 2002-2005. In this way we summarize the gathered information in order to obtain more complete data series and to avoid the effects of possible changes in measures and methods – something that often occurs when implementing surveys in their initial stages, such is the case of most of the entrepreneurship surveys. Averaging over a four years period seems not to cause a large bias in our estimations. We verified that there were not very important changes in scale or growth rate of the variables, and that aggregating information in other periodical base (e.g. three or five years) did not change drastically the outcomes. We found that the separation in these two periods is relatively stable and gave us a more complete data set.

4.2 Dealing with Missing Values

The *second approach* is the estimation of the missing values using the PPCA procedure suggested by Roweis (1998). Despite the averaging procedure previously discussed, available statistical series are not complete for all periods and all observations so our dataset may be plagued by problems of missing values. Various "solutions" have been applied in social sciences (Freudenberg, 2003) such as:

- dropping observations (data deletion), as in Cortinovic et al. (1993);
- replacing missing values with their means (mean substitution), as in Gwatkin et al. (2000), and Vyas and Kumaranayake (2006);
- replacing missing values with estimated or predicted values, i.e. using regressions based on other variables to estimate the missing values;
- multiple imputation using a large number of sequential regressions with indeterminate outcomes, which are run multiple times and averaged;
- nearest neighbor imputation, i.e. identifying and substituting the most similar case for the one with a missing value;
- simply, ignore them and take the average index of the remaining indicators within the component; and the list goes on.

Indicators (WDI), the International Monetary Fund (IMF)'s International Financial Statistics (IFS) and Government Finance Statistics (GFS), the OECD.Stats integrated system of databases for OECD countries and some benchmark economies, and United Nations' Statistical Database, are some well-known reliable sources of information that have in common the advantage that they gather and offer a lot of information that has been controlled for its reliability and comparability.

¹⁰This is the case of Ewing Marion Kauffman Foundation that collects information on entrepreneurial activity in the United States, at state-level. With this information they construct the *Kauffman Index of Entrepreneurial Activity* for the USA.

All of these methods have advantages and disadvantages. The advantages are generally associated with the simplicity or transparency of the procedure. The disadvantages are harder to deal with. For example, deleting observations that have missing data would significantly lower our sample size and the statistical power of our estimation would fall drastically. In fact, when constructing the entrepreneurship indicator, provided that it is more likely that developed countries have complete sets of statistics than developing countries, the indicators would be probably biased towards the relationships associated with the former countries, preventing us from learning about entrepreneurship in developing countries. On the other hand, replacing the missing values with the indicator’s mean will reduce variation among countries and increase the probability of clumping and truncation (see below). Substituting absent data with estimated or predicted values (regression approach) can introduce additional variability that is not related with the remaining indicators.

Nevertheless, statistical literature has recently centered its attention on the estimation of missing values conditional on available information for the rest of the sample in the context of principal components. Particularly appealing is the method called Probability PCA (PPCA) developed by Tipping and Bishop (1999). This model suggests that PCA can be considered a linear aggregation of Gaussian processes and that a maximum likelihood approach can be used to estimate its unknowns, namely the set of PCs. Roweis (1998) uses the Expectation-Maximization algorithm to obtain the PCs, and propose a simple extension to account for missing data (see the details in Appendix D). This is the statistical approach we adopt in this paper. Several simulations have been done to test out whether changes in the way the data set is completed has major impact on our CEI. Generally we have found that the outcomes vary little.

4.3 Outliers

Although in this work we do not deal with outliers, there are several methods to account for this problem. Chen (2002) makes a review of some of the most well-known methods to deal with outliers including

- Robust PCA by robustifying the covariance matrix;
- Robust PCA by Projection Pursuit;
- Robust PCA by Self-Organizing Neural Networks;
- Robust PCA by Weighted Singular Value Decomposition (SVD); and
- Torre and Black’s Algorithm.

Here we do not include the treatment of this problem, leaving it for future implementation of our entrepreneurship indicators.

4.4 Clumping and Truncation

As is discussed below, our weighting procedure is based on Principal Components Analysis (PCA) and a major challenge for PCA-based weighting is to ensure the range of variables is broad enough to avoid problems such as clumping and truncation. *Clumping* or *clustering* refers to the case in which countries present block-behavior, i.e. they behave as if they were put together in small clusters. This data characteristic generates strange behaviors of the variance-covariance matrix that is our basic piece of information.¹¹ *Truncation* implies that the observations are spread over a narrow range losing variability and biasing the outcomes towards more even distributions. We can detect these features through a good description of indicators (particularly, their ranges) and summary statistics (mean, standard deviation, interquartile range, maximum and minimum, etc.). One way to overcome these pitfalls is to add more variables to the analysis in order to capture full variability through other indicators. This is one of the reasons for using many variables to represent a certain category. Truncation may be very important in the case of developing countries given they are usually underrepresented in datasets. In our sample of 69 countries, they constitute about a half of the sample, in order to avoid the oversampling of developed countries.

¹¹As will be discussed below, we are working on the implementation of multi-level multi-block PCA that might improve the estimation of weights in presence of data clumping.

4.5 Relationship among indicators

To get a general idea on how the indicators relate to each other, and the appropriateness of their grouping, we analyze a simplified representation of the correlation matrix for the two sets of 65 indicators and 69 countries, distinguishing the grouping categories. Figure 2 shows the simplified correlations for the period 1998-2001, and Figure 3, for the period 2002-2005. As can be seen most of the significant correlations (more than 0.5 in absolute value) lie inside the shaded areas telling us that the grouping is at least reasonable. This does not mean that this is the only arrangement of indicators but it makes sense. Of course, we find several significant correlations outside the shaded areas, but they are generally weak. However the correlations outside the groups are consistent with the idea that all of these indicators have common behavior which we assume is due to an agglomerative concept, namely entrepreneurship.

5 Designing the Composite Indicator

A Composite Indicator (CI) is the mathematical combination of individual indicators that represent different dimensions of a concept whose description is the objective of the analysis (see Saisana and Tarantola, 2002). Composite Indicators are appealing due to their usefulness, flexibility, and simplicity. They constitute suitable tools for exploring less-known phenomena, and for benchmarking performances. Saisana and Tarantola (2002) highlight their usefulness to provide experts, stakeholders and decision-makers with:

- the direction of developments;
- comparison across places, situations and countries;
- assessment of state and trend in relation to goals and targets;
- early warning;
- identification of areas for action;
- anticipation of future conditions and trends; and
- communication channel for general public and decision-makers.

Constructing a CI is a good way to acquire knowledge at a relatively low cost provided CIs are also easily implemented, without requiring deep *a priori* knowledge or assumptions about the studied phenomenon, and generally they are not so computationally expensive. That is why they are extensively used in exploratory analysis, or for describing complex structures.

Our CEI shares all these properties and its weighting methodology accommodates a set of eight axioms (see Avanzini, 2008)¹². The highest hierarchical level of analysis is the *dimension* and indicates the scope of objectives, individual indicators and variables. Our indicator has seven dimensions (see Section 3). An *objective* indicates the direction of desired change, i.e. in what direction (upper or lower values) would an improvement in the indicator be reflected. *Individual Indicators* are the basis for evaluation in relation to a given objective (any objective may imply a number of different individual indicators). It is a function that associates each single country with a variable indicating its desirability according to expected consequences related to the same objective. A *Variable* is a constructed measure stemming from a process that represents, at a given point in space and time, a shared perception of a real-world state of affairs consistent with a given individual indicator¹³. In this context, the composite indicator or synthetic index is an aggregate or a function of all dimensions, objectives, individual indicators and variables used. This implies that what formally defines a composite indicator is the set of properties underlying its aggregation convention. The set of axioms ensures that our CEI will behave successfully for benchmarking and exploratory purposes.

¹²The weighting rule is built in order to incorporate the following axioms: Rationality, Weak Pareto Rule, Non-Dictatorship, Unrestricted Domain, Independence of Irrelevant Alternatives, Neutrality, Monotonicity, Reinforcement.

¹³Quoting Munda and Nardo (2005), "To give an example, in comparing two countries, inside the economic dimension, one objective can be 'maximization of economic growth'; the individual indicator might be R&D performance, the indicator score or variable can be 'number of patents per million of inhabitants'."

6 Estimating the Composite Indicator

From a practical perspective, it is very difficult to integrate individual indicators and variables in such a way that its outcome gives us a meaningful overview of the phenomenon, even more when we are interested in making inference based on that indicators. But given we have already defined the subject, its characteristics and how they can be captured through suitable variables, and we also have a theoretical framework that tell us which conditions the weighting procedure should accomplish (see Avanzini, 2008), we are ready to deal with the weighting problem itself, i.e. we are ready to construct the composite indicator.

We are interested in obtaining a weighting procedure with two additional characteristics beyond the axioms: first, we want the technique to be as *independent as possible from the analyst*, and second, that the procedure *extracts the maximum information* from available data. Taking into account these two characteristics we begin by discarding *ad-hoc* and subjective weightings, and concentrate in statistical and econometrical tools. In the literature many approaches have been used, most of them labeled under multivariate analysis methods, and aimed to extract information from several sort of data under different assumptions¹⁴.

In what follows we use Principal Component Analysis¹⁵ (PCA) to solve the weighting scheme. PCA is a powerful and relatively simple technique for extracting hidden structures from possibly high-dimensional datasets. Generally it is readily performed by solving an eigenvalue problem (standard PCA), or by using iterative algorithms which estimates the principal components, such as non-linear iterative partial least squares (NIPALS) and EM algorithm (Probabilistic PCA and Non-Linear PCA with EM algorithms). The intuition behind PCA is simple: suppose we have a dataset with a high number of variables (i.e. indicators) for various observations. One can think that these indicators are measuring the same object or episode from different perspectives so all of them contain common information about the object. PCA is an orthogonal transformation of the coordinate system in which we describe our data. The new coordinate values by which we represent the data are called *principal components*. It is often the case that a small number of such principal components is enough to account for the most of the structure in the data. These are sometimes called *factors* or *latent variables* of the data.

There are several ways for computing PCA depending on the underlying structure we have assumed. We concentrate in the comparison of the outcomes of four alternative PCA methods: standard PCA, NIPALS-based PCA, Multi-block PCA (Consensus PCA), and an extension of Probabilistic PCA with an EM algorithm for data reconstruction. Each method have particular characteristics that give rise to different set of weights, and of course, to different CEI. Here we develop four alternative weighting methods for estimating our CEI, discuss their strengths and weaknesses, and make reference to other relevant features.

6.1 Method (I): Overall Entrepreneurship Index

It consists of taking the whole set of variables (dataset) for each period and computing a standard PCA. There are two well-known variants for obtaining it: using covariance (with standardized indicators) or correlation matrix eigen-decomposition (see Jolliffe, 2002), or using iterative algorithms such as NIPALS (see Wold et al., 1987). The outcomes of both methods are the same though the eigen-decomposition approach is quicker. The square of the eigenvector associated to the first (bigger eigenvalue) principal component gives us the weighting matrix that multiplied by the dataset produces the weighted matrix of indicators, i.e. the CEI under this approach. We use this index as our benchmark and for the sake of comparability.

The indexes obtained with this approach are reported in Table 1 for the period 1998-2001, and in Table 5 for the period 2002-2005, and analyzed below (see next section). The theoretical background can be found in Appendix B.

6.2 Method (II): Disaggregated Entrepreneurship Index

This set of seven indexes per period, one for each dimension, is constructed using standard PCA on each group of standardized indicators. Each index ranks the country performance on the respective dimension,

¹⁴Just to mention some of them: discriminant analysis, correspondence analysis, multidimensional scaling, multivariate regression, etc.

¹⁵The origins of the method can be found in Pearson (1901), who coined the name and the mathematical principles; Hotelling (1933), that studied its properties; and Karhunen (1946). Recent literature reviews are contained in Diamantaras and Kung (1996) and Jolliffe (2002).

and is a useful tool to get a deeper insight on the driving forces behind entrepreneurship. Nonetheless, as with Method (I), it is difficult to assert the degree of entrepreneurship development provided that it is using only the information inside each dimension, so the remaining information (outside the shaded areas in Figures 2 and 3) is lost for the analysis. To get each index, we squared the eigenvector corresponding to the first principal component (largest eigenvalue) calculated over the set of indicators of each dimension solely, and then we multiply it by the dimensions' block of indicators to get a CI for each dimension.

The results for each dimension are shown in Table 2 (period 1998-2001) and Table 6 (period 2002-2005). Detailed explanations on the mathematical background for the PCA methodology used here can be found in Appendix B.

6.3 Method (III): Aggregated Entrepreneurship Index

Once we get the first principal component of each dimension, and the associated index from Method (II), a natural extension might be to group the seven principal components (i.e. we accommodate the PCs containing the maximum common information of each dimension) in a single matrix. Each PC represents an important proportion of the information contained in that dimension, so the matrix containing such PCs is our best set of common knowledge about the behavior of the entrepreneurship dimensions, constrained to use just one PC per dimension. Additionally, PCs are standardized random variables given they are constructed as a weighted sum of standardized random variables. If we then apply standard PCA to this matrix we would get the Aggregated Entrepreneurship Index. This composite indicator is a summary of the most relevant information of each dimension.

Although we have gathered together all the dimension-related information, much of the interactions among variables of different dimensions is only captured through their impact on the dimensional PCs: indicators that are weak within a dimension, but highly correlated with other indicators in the remaining dimensions, could be underweighted in the aggregate index¹⁶. Again the explanations related to the mathematical procedure are given in Appendix B, and the indexes are shown in Table 2 (period 1998-2001) and Table 6 (period 2002-2005).

6.4 Method (IV): Multi-Dimensional Entrepreneurship Index

Finally we arrive to the key proposal of this paper. We would like to get a composite indicator capable of accounting for particular intra-dimensional structures, but also inter-dimensional relationships, and overall behavioral patterns. The Consensus PCA (CPCA) is a method that can account for all these problems and still gives us a consistent, meaningful set of weights¹⁷. The intuition behind CPCA is explained in Wold et al. (1987) in the context of the analysis of sensory data: suppose that a certain number of referees are judging the sensory quality of a number of wines. Each referee gives his judgement on various quality characteristics of the wines (body, bitterness, color, etc.). The results on all the tests for each judge are placed in separate blocks conforming the sub-level. The *summary scores* of each judge are the *block scores*, and the consensus of all judges is represented by a *super score* summarizing individual scores. The weights given to each block in this *super-level* show the relative importance of each judge in the *consensus score* (for details on its computation see Appendix C).

One of the advantages of CPCA is that it uses all the available information in each iteration: it weights block indicators to get the block scores (in the spirit of Method II), then it uses these scores to construct a super-level block of information and estimates its score (in the spirit of Method III), and finally it re-weights the sub-level indicators with the overall weights (in the spirit of Method I), and iterates this procedure until convergence. Two additional advantages of CPCA are: (i) it has the same objective function as standard PCA, i.e. maximization of the variance of the dataset; and (ii) it automatically adjust relative variance of different indicators by virtue of its block scaling procedure (Westerhuis et al., 1998). Also notice that Method (IV) overcomes the drawbacks posed for each of the other methods. First, Method (I) was not able to focus in certain areas of the correlation matrix, those related with the dimension blocks, and uses all the available

¹⁶In a certain way we are violating the Reinforcement Axiom. See Avanzini (2008) on this point.

¹⁷Borrowed from chemical and batch processes monitoring, this method is based in non-linear iterative partial least squares (NIPALS) and it was introduced by Wold et al. (1987). As was previously discussed, NIPALS is one of the alternatives to estimate PCs by means of an iterative weighting algorithm. This weighting algorithm allows the differentiation of multiple layers of information aggregation through successive re-weighting of indicators under a hierarchical categorization.

information without discriminating. Second, Method (II) accounted for each dimension behavior but was unable to integrate the information remaining out of the shaded areas in Figures 2 and 3. Finally, Method (III) captured variability at dimensional level, and used the reduced information dataset to estimate the relationship among dimensions. However, there were no possibility to capture more complex relationships (possibly represented by higher order principal components). The three drawbacks are elegantly managed in the context of CPCA through the NIPALS algorithm that allows the integration of dimensional (block) and overall information through an iterative procedure.

The set of weights (at dimensional level and at variable/indicator level) are shown in Tables 3 and 7, and indexes are shown in Tables 4 and 8, for period 1998-2001 and 2002-2005, respectively. Appendix C develops the mathematical background.

6.5 Some Comments on the Estimation Strategy

Although we have made a big effort to rely as much as possible upon statistical methods to construct our CEI, we have been forced to use our own judgement repeatedly. Provided the outcomes of the CEI depend largely on the selected approach, we conducted sensitivity tests to analyze the impact of including or excluding variables, changing the weighting scheme, using different standardization techniques, selecting alternative base years, excluding cases with unreliable data, etc., on the results of the CEI. We used bootstrapping methods to account for these issues, and generally we found no significant (95% confidence level) variations in the outcomes.

We also realize that the quality of our dataset prevent us from doing categorical assertions: averaging over periods makes us lose some of the dynamics of the entrepreneurship process; the methodology we adopt to deal with missing data –though being quite reliable– has been forced to deal with many missing values, weakening the outcomes¹⁸. Although the dataset we finally use seems to be quite stable and reliable, we know that countries showing originally more complete datasets are those more developed, with higher per capita income, with better life standards, with higher mean levels of education, etc., while developing countries have more incomplete datasets, so the outcomes referring to developed countries might be more robust than those involving developing countries.

Now we turn back to our CEI. We start stating that our idea of grouping the set of variables in nine categories, finally reduced to 7 to robustify the estimation of block scores, is a good approximation for the set of available data. Indeed, Figure 4 shows the Scree plots, i.e. the plot of the eigenvalues, and is hard to detect such an "elbow" as is suggested under this approach. Using estimations of the intrinsic dimensionality (van der Maaten, 2007) we find that for bootstrapped samples, the dimensions roughly vary among 8 and 10. So our categorization (as showed in Tables 3 and 7) are a good grouping strategy. Moreover, observing the correlation patterns in Figures 2 and 3, it seems to be the case that the most significant relationships between indicators are located inside the shaded areas that correspond to the correlation matrixes of each block variables. However, to get more robust results we regroup the first three categories (Firm Dynamics, Firm Survival, and Ownership) in a single category to strengthen block scores estimation. We perform some simulations to evaluate how the results change, and the final distribution of indexes seems to remain unchanged with this procedure.

7 Analyzing Entrepreneurship using the CEI

The Overall Entrepreneurship Index, our benchmark technique, shows that less developed countries improve their performance in the second period. The indexes are reported in Table 1 for the period 1998-2001, and in Table 5 for the period 2002-2005. They are calculated over standardized variables so their means are zeroes, and the standard deviations are about 0.30 for 1998-2001, and 0.22 for 2002-2005. The range of the indexes narrows from 1998-2001 to 2002-2005 showing that countries are making efforts to improve the entrepreneurial activity. Those efforts narrow entrepreneurial gaps among countries. Little can be said about the index itself: as was discussed, the way the index weights all the attributes prevent us from using

¹⁸Roweis (1998) tested out the algorithm with a dataset that have one missing dimension on each observation, and he considered the algorithm performed successfully under that constraint. Here, we are forcing it to deal with many missing indicators for every observation.

it to infer which are the driving forces behind entrepreneurial development, and in what areas it has major impact. Nevertheless we can still exploit it as a benchmarking tool that gives us some information on the way different countries are performing. But we necessarily need to compare the behavior of each indicator to know more about the reasons of these behavior. This index, even though simple, is restrictive for inference purposes, and is difficult to get more information than the ranking of countries.

The second CEI, the Disaggregated Entrepreneurship Index, explores the behavior of each of the seven categories. They are reported in Tables 2 and 6 for periods 1998-2001 and 2002-2005, respectively. The behavior is more erratic at this level of aggregation. Results are difficult to interpret: for example, in 2002-2005, the disconnection between Knowledge Procurement and Innovation (something that can be seen in the correlation matrix, Figure 3) produces rankings under these categories that are very different¹⁹. On the side of the ease of entrepreneurial activity, at the end of the '90s, Japan and the USA appeared as the most opened economies to this sort of activities. But with Brazil, Paraguay, Bolivia and other developing countries entering the international scene, the classically entrepreneurial countries were relegated to secondary places. Entrepreneurial vocation is stronger in countries such as New Zealand and Australia, but less developed economies like Mexico and Brazil appear to have the necessary intention to get more involved in entrepreneurship development. This is consistent with the evidence that GEM's TEA index shows when it is drawn against GDP per capita: less developed countries are more likely to develop entrepreneurial activities, and are well disposed towards entrepreneurship than more developed countries. But the driving forces are different for both groups of countries: while less developed countries are performing entrepreneurial activities by *necessity* (generally associated with high rate of unemployment in the formal sector), high-income economies focus on *opportunity entrepreneurship*, mostly linked to product innovation, new markets, and improved production processes. The *necessity entrepreneurship* is more related to "economic survival". Within our dataset, Mexico and Brazil are good examples of this situation, as can be inferred from the impact of entrepreneurship in the Employment dimension.

When we aggregate the seven dimensions on a single CEI, the Aggregated Entrepreneurship Index (first ranking in Tables 2 and 6), we find that the USA, China and Japan are consolidated as leading entrepreneurial economies. Nonetheless, the indexes are very different between them, with abrupt changes in rankings that are not so convincing. The Employment dimension have a strong weight (see the weights on the top of each dimension) and it is dominant in the 2002-2005 period. But differences in the range of other less relevant dimensions contribute to generate some important distortions, for example, placing Finland in the 63th in 2002-2005 after being in the 10th position in 1998-2001.

An interesting feature of this CEI, is the way it weights different dimensions²⁰: in 1998-2001, the dimensions were weighted in three ranges: less than 5%, between 11% and 13% and beyond 20%. In 2002-2005, the dimensions Employment and Economic Activity account for almost 66% of the total variability being the leading ones. Surprisingly, Knowledge Procurement and Innovation only account for 20% in 2002-2005 while in 1998-2001 those dimensions constituted about 55% of the variability. Also the weak role of Entrepreneurial Activity in 2002-2005 is suspicious. With the current dataset, this weighting process seems to be very unstable and its outcome is not very sounding. However we have to recognize that during 2002-2005 many developing economies presented amazing economic growth rates and improvements in their employment, both phenomena captured by an increased weighting of Employment and Economic Activity dimensions.

Beyond some particular problems arising probably due to data quality, from a methodological perspective, one of the advantages of the composite index obtained in this way is that it is able to capture the areas or dimensions where there is still "action", in the sense that the highly-weighted dimensions are the most relevant differences between countries, and are the areas where major improvements can be done.

Finally we turn our attention to the Multidimensional Entrepreneurship Index that is our starred CEI. Tables 3 and 7 report the relative weighting of variables and dimensions, and Tables 4 and 8 report scores and rankings. The novelty of this index, as was previously discussed, is its capability to capture information inside and outside the specific dimension, through an iterative weighting process that gives us deeper insight on entrepreneurship behavior. The driving forces in 1998-2001 (Table 3) were related with economic activity (particularly the creation of small and medium enterprises). Entrepreneurship spirit and business

¹⁹ As an example: Indonesia is assigned the third place under the Innovation dimension, and the last place under the Knowledge Procurement dimension.

²⁰ Each dimension's weight can be found in the top of the dimension's disaggregate index. These weights correspond to the importance each dimension has in the aggregate index.

environment were almost equally weighted as were knowledge procurement and innovation dimensions. The number of technicians in R&D, and the submission of patent applications and trademarks were relevant aspects of the entrepreneurial development. Also domestic credit played an important role in the strengthen of the entrepreneurial activity.

In 2002-2005 (Table 7), countries have improved those dimensions and they have lost their importance. Innovation is now more important and the contribution of entrepreneurial development to employment appears as the major impact. Barriers to entrepreneurship and a suitable environment for business developments have lost their importance given that most of the countries made major reforms and efforts to support entrepreneurial activity. Again the number of technicians in R&D, and the submission of patent applications and trademarks have been important determinants of entrepreneurial performance. The impact on economic activity has been diminished to highlight the employment dimension instead.

In both periods, the dimension reflecting entrepreneurial activity itself has not been important. We find that this is consistent with the fact that information related with firm dynamics and ownership is more scarce as is less represented in the dataset than other information. Perhaps in the future, with improved datasets, this dimension would become more dominant in the determination of the index value. With our current dataset, many of the characteristics measured under Entrepreneurial Activity have been measured under different dimensions with proxies that present a more complete record for the sample.

Developing countries have generally improved their positions in the rankings. Many reforms were conducted at the end of the '90s and during the first years of this century. Simplified taxes schemes, easier business registration, and more available financial information (that speed the credit market up) contributed to the generation of a supporting system for the entrepreneurial activity. Also the economic prosperity driven by the boom of commodity prices generated a market for new small and medium enterprises that hired new employees. China, Mexico, Brazil, Indonesia, and Argentina are good examples of high growth rates that were accompanied by new employment, new small and medium enterprises, and technological development. These improvements in the economy and particularly in the entrepreneurial sector in developing countries do not mean that developed countries have worsened their policies: it only means that developing countries have made changes at a quicker rhythm, and the impact of their reforms and economic changes have had major impact on the economy than it was the case in developed countries.

Recall that the most important characteristic of this index is the way it selects the determinants of entrepreneurship, providing us with a closer insight into how policies and efforts should be oriented. And this is the central message of this approach: the suggested methodology helps us to direct policies towards weak determinants to improve them, and towards strong impacts to ensure they will be maintained or improved. This approach to the estimation of the CEI has the following advantages that make it an "ideal" device (among the ones we know) for entrepreneurship analysis:

- it provides us with the traditional benchmarking tool: all countries are compared on the same bases;
- it allows us to rank countries' performance;
- it informs about intensity: the index tells us up to what degree countries differ in entrepreneurial development;
- and finally, it gains an insight into the importance of entrepreneurship determinants and impacts.

8 Concluding Remarks

However much relevant issues have been pointed out, a lot of work has to be done: improving data collection and its reliability is the first step. Continue exploring robust methodologies to account for variations in entrepreneurship performance is another task, that in some way we tried to tackle here, but it needs closer attention. Finally, researchers should be more involved in determining the driving forces behind entrepreneurship and how those ones relate to economic growth and welfare, while policy- and decision-makers should be more opened to the possibility of changing policy directions to account for this arising phenomenon.

A communicational issue needs to be addressed: it might be very tempting for politicians to show how countries satisfactorily evolve in international benchmarks, but as was explained previously, the way the CEI reports changes (specially that based on Consensus PCA) makes it less attractive because when all countries

get the message and improve their policies in a certain direction, differences arising from that dimension are not still relevant for the problem, and politicians might feel that being equated to other countries is perhaps not enough to capture voters attention. We recognize that this can be a pitfall of this CEI though its ease of communication.

We have to be aware of the way we should use the tools explained in this document: the CEIs tell us where we stand in the global economy (benchmarking purpose), and inform us which might be the principal reasons of our situation and how far are we from the best performers (inference purpose), allowing policy-makers to change policies direction to support entrepreneurship (policy purpose). However, weak datasets or careless use of the methodologies discussed in this paper may result in misleading interpretations and perverse policies that might block rather than encourage entrepreneurial activity. Here we discussed some of the problems with datasets and suggested ways of fixing them. We also discussed the problems of inference and interpretation of results. But as always, the art of dealing with an intricate subject such as entrepreneurship is in the hands of the analyst.

References

- [1] Ahmad, Nadim, and Anders N. Hoffmann (2008). "A framework for addressing and measuring Entrepreneurship", OECD Statistics Directorate, Working Paper STD/DOC(2008)2, January.
- [2] Ahmad, Nadim, and Richard G. Seymour (2008). "Defining Entrepreneurial Activity: Definitions supporting frameworks for data collection", OECD Statistics Directorate, Working Paper STD/DOC(2008)1, January.
- [3] Audretsch D. B. and R. Thurik (2001): "Linking Entrepreneurship to Growth", STI Working Paper 2001/2, OECD, Paris.
- [4] Audretsch, D. B. (2002). "Entrepreneurship: A Survey of the Literature", Prepared for the European Commission, Enterprise Directorate General, 2002.
- [5] Avanzini, Diego B. (2008). "Composite Entrepreneurship Indicators: theoretical derivation and general methodology", unpublished manuscript, Catholic University of Chile.
- [6] Bartholomew, D. J., F. Steele, I. Moustaki and J. I. Galbraith (2002). "The analysis and interpretation of multivariate data for social scientists", Chapman & Hall/CRC.
- [7] Behrens, Axel (2007). "Entrepreneurship Indicators: performance, determinants and impact", European Commission, presented at OECD/Eurostat 2007, May.
- [8] Cantillon, Richard (1755, re-edited 1931). *Essai sur la Nature du Commerce en Général*. London, UK: MacMillan.
- [9] Cattell, R. B. (1966). "The Scree test for the number of factors", *Multivariate Behavioral Research*, vol. 1, pp. 245-276.
- [10] Chen, Haifeng (2002). "Principal Component Analysis with Missing Data and Outliers", unpublished manuscript, Electrical and Computer Engineering Department, Rutgers University.
- [11] Commission of the European Communities (2003). *Green Paper: Entrepreneurship in Europe*. In *Enterprise*, editor: Enterprise Publications.
- [12] Cortinovis, I., V. Vela and J. Ndiku (1993). "Construction of a socio-economic index to facilitate analysis of health in data in developing countries", *Social Science and Medicine*, vol. 36, pp. 1087-97.
- [13] Davis, Tim (2007). "Entrepreneurship Indicators Project: Developing comparable measures of entrepreneurship and the factors that enhance or impede it", presentation at the World Forum on Statistics Knowledge and Policy Workshop on Entrepreneurship Indicators, OECD Statistics Directorate, June 27, Istanbul, Turkey.

- [14] Diamantaras, K. I. and S.-Y. Kung (1996). *Principal Component Neural Networks. Theory and Applications*, Wiley Series on Adaptive and Learning Systems for Signal Processing, Communications, and Control, NY.
- [15] Drucker, Peter F. (1985). *Innovation and Entrepreneurship: Practice and Principles*. New York, USA: HarperBusiness.
- [16] Eastment, H. T. and W. J. Krzanowski (1982). "Cross-validators choice of the number of components from a principal component analysis", *Technometrics*, vol. 24, pp. 73-77.
- [17] Farmer, S. A. (1971). "An investigation into the results of principal component analysis of data derived from random numbers", *Statistician*, vol. 20, pp. 63-72.
- [18] Filmer, D. and L. H. Pritchett (2001). "Estimating wealth effects without expenditure data - or tears: an application to educational enrollments in states of India", *Demography*, vol. 38, pp. 115-132.
- [19] Freudenberg, Michael (2003). "Composite indicators of country performance: a critical assessment", STI Working Paper 2003/16, Industry Issues, OECD.
- [20] Fukuda, Shin-ichi, and Takashi Onodera (2001). "A New Composite Index of Coincident Economic Indicators in Japan: How Can We Improve the Forecast Performance?", Discussion Paper, CIRJE-F-101, January.
- [21] Gallup Europe (2002). Flash Eurobarometer 124, EU Enlargement, UK Opinion Poll, EOS, April. Accessed at http://europa.eu.int/comm/public_opinion/index_en.htm
- [22] Giri, N. C. (2004). *Multivariate statistical analysis*. New York: Marcel Dekker Inc.
- [23] Guttman, L. (1954). "Some necessary conditions for common factor analysis", *Psychometrika*, vol. XIX, pp. 149-161.
- [24] Gwatkin, D. R., S. Rustein, K. Johnson, R. P. Pande and A. Wagstaff (2000). *Socio-economic differences in health, nutrition and population in Brazil, Ethiopia, Nigeria, and Vietnam*. Washington, DC: HNP/Poverty Thematic Group of the World Bank.
- [25] Hotelling, H. (1933). "Analysis of a complex of statistical variables into principal components", *Journal of Educational Psychology*, vol. 24, pp. 417-441.
- [26] Houweling, T. A. J., A. E. Kunst and J. P. Mackenbach (2003). "Measuring health inequality among children in developing countries: does the choice of the indicator of economic status matter?", *International Journal for Equity in Health*, vol. 2, pp. 8-19.
- [27] Hung, Wendy S. M. (2003). "Experimental Composite Leading Indicator for Hong Kong, China", Census and Statistics Department, Hong Kong, China. Paper presented at OECD/ESCAP Workshop on Composite Leading Indicators and Business Tendency Surveys, Bangkok, 24 - 26 February.
- [28] Ireland, R. D., M. A. Hitt and D. G. Sirmon (2003). "A Model of Strategic Entrepreneurship: The Construct and its Dimensions", *Journal of Management*, vol. 29, No. 6, pp. 963-989.
- [29] Jackson, D. A. (1993). "Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches", *Ecology*, vol. 74, pp. 2204-2214.
- [30] Jobson, J. D. (1992). *Applied multivariate data analysis*. New York: Springer-Verlag.
- [31] Jolliffe, I. T. (1972). "Discarding variables in a principal component analysis I: Artificial data", *Applied Statistics*, vol. 21, pp. 160-173.
- [32] Jolliffe, I. T. (1973). "Discarding variables in a principal component analysis II: Real data", *Applied Statistics*, vol. 22, pp. 21-31.
- [33] Jolliffe, I. T. (1987). "Comment: Selection of variables", *Applied Statistics*, vol. 36, pp. 373-374.

- [34] Jolliffe, I. T. (2002). *Principal Component Analysis*. Second Edition. New York: Springer-Verlag.
- [35] Kaiser, H. F. (1960). "The application of electronic computers to factor analysis", *Educational and Psychological Measurement*, vol. 20, pp. 141-151.
- [36] Kaiser, H. F. (1961). "A note on Guttman's lower bound for the number of common factors", *British Journal of Psychology*, vol. 14, pp. 1-2.
- [37] Kaplan, R. S. and D. P. Norton (1992). "The balanced scorecard: measures that drive performance", *Harvard Business Review*, January-February, pp. 71-80.
- [38] Kaplan, R. S. and D. P. Norton (1993). "Putting the Balanced Scorecard to Work", *Harvard Business Review*, September-October, pp. 2-16.
- [39] Kaplan, R. S. and D. P. Norton (1995). "Chemical Bank: Implementing the Balanced Scorecard" Harvard Business School Press.
- [40] Kaplan, R. S. and D. P. Norton (1996a). "Using the balanced scorecard as a strategic management system", *Harvard Business Review*, January-February, pp. 75-85.
- [41] Kaplan, R. S. and D. P. Norton (1996b). "Balanced Scorecard: Translating Strategy into Action" Harvard Business School Press.
- [42] Karhunen, K. (1946). "Zur spektraltheorie stochastischer prozesse", *Annales Academiae Scientiarum Fennicae, Series A1, Mathematica-Physica*, vol. 34, pp. 1-7.
- [43] Kirzner, I. M. (1973). *Competition and Entrepreneurship*. Chicago, Illinois: The University of Chicago Press.
- [44] Knight, Frank. (1921, 1st. edition / 1971). *Risk, Uncertainty And Profit*. Chicago: University of Chicago Press.
- [45] Legendre, L. and P. Legendre (1983). *Numerical Ecology*. Amsterdam: Elsevier.
- [46] Leitão da Silva Martins, Susana Paula (2007). "Indicators for measuring entrepreneurship: a proposal for a scoreboard", *Industry & Higher Education*, pp. 85-97, February.
- [47] Lumpkin, G. T. and G. G. Dess (1996). "Clarifying The Entrepreneurial Orientation Construct and Linking It to Performance", *The Academy of Management Review*, vol. 21, No. 1, pp. 135-172.
- [48] Manly, B. F. J. (1994). *Multivariate statistical methods. A primer*. Second Edition. London: Chapman and Hall.
- [49] Marshall, Alfred (1890). *Principles of Economics*. Reprinted in 1997. London, UK: Prometheus.
- [50] Munda, Giuseppe, and Michela Nardo (2005). "Non-compensatory composite indicators for ranking countries: a defensible setting", *Institute for the Protection and Security of the Citizen, European Commission, EUR 21833 EN*.
- [51] Nilsson, Ronny (1987). "OECD Leading Indicators", *Economic Studies*, No. 9.
- [52] Nilsson, Ronny (2000). "Confidence Indicators and Composite Indicators", *OECD Statistics Directorate, paper presented at the CIRET Conference, Paris, 10-14 October*.
- [53] Nilsson, Ronny (2003). "OECD System of Leading Indicators: Practices and Tools", *OECD Statistics Directorate. Paper presented at OECD/ESCAP Workshop on Composite Leading Indicators and Business Tendency Surveys, Bangkok, 24 - 26 February*.
- [54] OECD (1998). "OECD Composite Leading Indicators: a tool for short-term analysis", *OECD, Statistics Directorate, November*.

- [55] OECD (2006). "Understanding entrepreneurship= developing indicators for international comparisons and assessments", OECD Statistics Directorate, Committee on Statistics, Report on the OECD's Entrepreneurship Indicators Project and Action Plan, STD/CSTAT(2006)9, May 24th.
- [56] OECD/Eurostat (2007). "OECD/Eurostat Entrepreneurship Indicators Programme, Indicator Framework", definitive version, October.
- [57] Pearson, Karl (1901). "On lines and planes of closestfit to systems of points inspace", The London, Edinburgh and Dublin Philosophical Magazine and Journal of Sciences, vol. 6, pp. 559–572.
- [58] Penrose, E. T. (1959 / re-edited 1980). *The Theory of the Growth of the Firm*. Oxford, UK: Basil Blackwell.
- [59] Petit, Gérald, Gérard Salou, Pierre Beziz and Christophe Degain (1996). "An update of OECD Leading Indicators", paper presented at the OECD Meeting on Leading Indicators, Paris, October.
- [60] Reynolds, P. (1991). "Sociology and Entrepreneurship: Concepts and Contributions", *Entrepreneurship Theory and Practice*, vol. 16, No. 2, pp. 47-70.
- [61] Roweis, Sam (1998). "EM algorithms for PCA and SPCA," in *Advances in Neural Information Processing Systems*, vol. 10. MIT press, pp. 626–632.
- [62] Saisana, Michaela and Stefano Tarantola (2002). "State-of-the-art report on current methodologies and practices for composite indicator development," EUR 20408 EN, European Commission-JRC: Italy.
- [63] Say, Jean-Baptiste (1803 / re-edited 1964). *A Treatise on Political Economy; or the Production, Distribution and Consumption of Wealth*. New York: Augustus M. Kelley ed.
- [64] Schumpeter, J. A. (1934). *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*. Cambridge, USA: Harvard University Press.
- [65] Schumpeter, J. A. (1933). "The common sense of econometrics", *Econometrica*, vol. 1, pp 5-12.
- [66] Schumpeter, J. A. (1942, edited 1949). *Capitalism, Socialism and Democracy*. Harper & Row, New York, NY.
- [67] SEIFA (2001). "Census of Population and Housing: Socio-economic Indexes for Area's (SEIFA)", Technical paper, Australian Bureau of Statistics, 2039.0.55.001.
- [68] SEIFA (2006). "Socio-economic Indexes for Areas (SEIFA) - Technical Paper", Australian Bureau of Statistics, 2039.0.55.001.
- [69] Shane, S. and S. Venkataraman (2000). "The Promise of Entrepreneurship as a Field of Research", *The Academy of Management Review*, vol. 25, No. 1, pp. 217-226.
- [70] Shaver, K. G. and L. R. Scott (1991). "Person, process, choice: The psychology of new venture creation", *Entrepreneurship Theory and Practice*, Vol. 16, No. 2, pp. 23-45.
- [71] Statistics Netherlands (2007). "Figures on the investment climate in the Netherlands 2007", Report, Voorburg/Heerlen.
- [72] Stevenson, H. (1985). "The Heart of Entrepreneurship," *Harvard Business Review*, March-April, pp. 85-94.
- [73] Thorton, P. (1999). "The Sociology of Entrepreneurship" *Annual Review of Sociology*, vol. 25, pp. 19-46.
- [74] Tipping, M. E. and C. M. Bishop (1999). "Mixtures of probabilistic principal component analyzers," *Neural Computation*, vol. 11, no. 2, pp. 443–482.
- [75] United Nations (2001). "Indicators of sustainable development: Guidelines and methodologies", UN Division for Sustainable Development.

- [76] VaLUENTIS (2006). Human Capital Composite Index, UK Local Government Rankings. Serie 'Cracking the Human Capital Code study", Volume IIA, September 30.
- [77] van den Berg, Frans (2001). Multi-Block Toolbox for Matlab, version 0.1 June 06, 2001, KVL, Denmark: <http://www.models.kvl.dk/source/MBtoolbox>
- [78] van der Maaten, L. J. P. (2007). "An Introduction to Dimensionality Reduction Using Matlab: ", Universiteit Maastricht MICC/IKAT, Report MICC 07-07, The Netherlands, July.
- [79] Verbeek, Jakob (2006). Probabilistic PCA with missing values, Matlab code: http://lear.inrialpes.fr/~verbeek/code/ppca_mv.m
- [80] von Mises, L. (1949 / re-edited 1996). Human Action: A Treatise on Economics. Fourth Revised Edition. San Francisco: Fox & Wilkes.
- [81] Vyas, Seema and Lilani Kumaranayake (2006). "Constructing socio-economic status indices: how to use principal components analysis", Health Policy and Planning, vol. 21, pp. 459–468.
- [82] Walras, L. 1954. Elements of Pure Economics, or The Theory of Social Welfare. London: Allen & Unwin for the American Economic Association and the Royal Economic Society.
- [83] Westerhuis, J. A., Th. Kourti and J. F. MacGregor (1998). "Analysis of Multiblock and hierarchical PCA and PLS Models", Journal of Chemometrics, vol. 12, pp. 301-321.
- [84] Wold, S. (1978). "Cross-validators estimation of the number of components in factor and principal component models", Technometrics, vol. 20, pp. 397-405.
- [85] Wold, S., K. Esbensen and P. Geladi (1987). "Principal Component Analysis", Chemometrics and Intelligent Laboratory Systems, vol. 2, pp. 37-52.
- [86] Yeomans, Keith A. and Paul A. Golder (1982). "The Guttman-Kaiser Criterion as a predictor of the number of common factors", The Statistician, vol. 31, No. 3, September, pp. 221-229.
- [87] Yu, Chang-Ching, John T. Quinn, Christian M. Dufournaud, Joseph J. Harrington, Peter P. Rogers and Bindu N. Lohani (1998). "Effective dimensionality of environmental indicators: a principal component analysis with bootstrap confidence intervals", Journal of Environmental Management, Vol. 53, Issue 1, pp. 101-119.
- [88] Zabala-Iturriagoitia, Jon Mikel, Fernando Jiménez-Sáez, Elena Castro-Martínez and Antonio Gutiérrez-García (2007). "What indicators do (or do not) tell us about Regional Innovation Systems", Scientometrics, Vol. 70, No. 1, pp. 85-106.

A Data Sources and Variables Description

Following is a list of the used indicators and variables, are they are listed in Tables 4 and 10, gathered under each data source.

A.1 From COMPENDIA:

COMPENDIA (Comparative Entrepreneurship Data for International Analysis) is an across-countries and -time comparable dataset for 23 developed countries covering information from 1970 onwards. The data series taken from this database enter Table 4 as indicators 7-9, and Table 10 as indicators 9-11. The indicators are:

Business Ownership Rates: This is the number of business owners divided by total labour force. Only persons who are self-employed as their main occupation are included in the figures. The owners are classified depending on the sector they develop their entrepreneurial activity: private or public sector, and in private sector, agricultural and non-agricultural sub-sectors.

A.2 From GEM:

The New Entrepreneurship International (GEM) dataset contains annual harmonized data on early-stage entrepreneurial activity for 43 countries since 1998. The abbreviation GEM stands for Global Entrepreneurship Monitor and is the common name for this international survey. The data series taken from this database enter Table 4 with indicators 5, 6, 29-32, 34, 35, 37, and 38, and Table 10 with indicators 6-8, 34-39, and 41-43. All the indicators are taken relative to the adult population 18-64 years. The selected indicators are:

Total Entrepreneurial Activity (TEA) Index: number of people currently setting up a business or owning/managing a business existing up to 3,5 years.

Necessity Entrepreneurial Activity Index: number of people involved in entrepreneurial activity (TEA) out of necessity.

Opportunity Entrepreneurial Activity Index: Measures the number of people involved in entrepreneurial activity (TEA) out of opportunity.

Female Total Entrepreneurial Activity Index: Measures the number of women involved in entrepreneurial activity (TEA).

Nascent Entrepreneurial Activity Index: Measures the number of people currently setting up a business.

Young Firm Entrepreneurial Activity Index: Measures the number of people owning/managing a business that exists up to 3,5 years.

Established Businesses Activity Index: Measures the number of people owning/managing a business that exists over 3,5 years.

Future Entrepreneur Index: Share of people expecting to start a business within three years.

Know Entrepreneur Index: Share of people that personally know someone who started a business in the past two years.

Potential Entrepreneur Index: Share of people indicating to have the required skills and knowledge for setting up a business themselves.

Fear of Failure Index: Share of people that would abstain from setting up a business when they would sense a fear of failure.

Informal Investors Index: Measures the number of people investing own money to start-ups.

A.3 From EIM:

The dataset International Benchmark of Entrepreneurs from Entrepreneurship International Monitor (EIM) contains data about firm entries, firm exits and bankruptcies. Therefore, 9 countries from the EU and additionally the USA and Japan, are included in this set. The figures in this set are comparable across countries and over time. Additionally, the set has data of fast growing firms with a high employment growth and/or a high sales growth, measured for periods of three years. The performances of these companies are compared to companies with an 'average growth pattern'. The indicators 1-4, 10-17, and 19-24 in Table 4, and 1-4, 13-21, and 24-29 in Table 10 are based on the following:

Entry rate: number of new 'activities' started by entrepreneurs, divided by the total number of companies in a certain country.

Exit rate: number of 'activities' that finished their activities, divided by the total number of companies in a certain country.

Average size of firm entries: average number of workers in the new 'activities'.

Share of entries in employment: number of new 'activities', divided by the total number of workers in a certain country.

Share of exits in Employment: number of companies that stopped their activities, divided by the total number of workers in a certain country.

Bankruptcy rate: number of bankruptcies, divided by the total number of companies in a certain country.

Share of bankruptcies in firm exits: number of bankruptcies, divided by the total number of companies that stopped their activities.

Average sales for all enterprises - last year of period: for a certain 3-years period, average sales (in regard to companies) for the last year.

Average sales for all enterprises - growth rate over whole period: for a certain 3-years period, growth rate of the average sales over the whole period.

Average sales for not fast growing enterprises - last year of period: for a certain 3-years period, average sales (in regard to not-fast growing companies) for the last year.

Average sales for not fast growing enterprises - growth rate over whole period: for a certain 3-years period, growth rate of the average sales over the whole period (not-fast growing companies).

Average sales for fast growing enterprises - last year of period: for a certain 3-years period, average sales (fast growing companies) for the last year.

Average sales for fast growing enterprises - growth rate over whole period: for a certain 3-years period, growth rate of the average sales over the whole period (fast growing companies).

Average number of workers for all enterprises - last year of period: for a certain 3-years period, average number of employees (in regard to companies) for the last year.

Average number of workers for all enterprises - growth rate over whole period: for a certain 3-years period, growth rate of the number of employees over the whole period.

Average number of workers for not-fast growing enterprises - last year of period: for a certain 3-years period, average number of employees (not-fast growing companies) for the last year.

Average number of workers for not fast growing enterprises - growth rate over whole period: for a certain 3-years period, growth rate of the average number of employees over the whole period (not-fast growing companies).

Average number of workers for fast growing enterprises - last year of period: for a certain 3-years period, average number of employees (fast growing companies) for the last year.

Average number of workers for fast growing enterprises - growth rate over whole period: for a certain 3-years period, growth rate of the average number of employees over the whole period (fast growing companies).

A.4 From Freudenberg (2003):

The dataset covers innovation- and knowledge-procurement-related indicators for the last years of the '90s. The indicators enter Table 4 with numbers 47-54, and 58-61. Their description follows:

R&D performed by the non-business sector as a percentage of GDP: is a proxy for a country's relative efforts to create new knowledge, though it should be noted that new knowledge can also originate in firms or in partnership with firms. Public R&D activities also disseminate new knowledge and exploit existing knowledge bases in the public sector. It is reported as a ratio for 1995-99.

Number of non-business researchers per 10,000 labour force: (self-explanatory).

Expenditures on basic research as a percentage of GDP: Basic research is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundation of phenomena and observable facts, without any particular application or use in view.

Ratio of science, engineering and health PhDs per population aged 25 to 34 years: This age group was chosen because it is the only one for which there are internationally comparable data.

Number of scientific and technical articles per million population: article counts of scientific research are based on scientific and engineering articles published in approximately 5.000 of the world's leading scientific and technical journals.

Business-financed R&D performed by government or higher education as a percentage of GDP: R&D expenditure financed by industry but performed by public research institutions or universities.

Number of scientific papers cited in US-issued patents per million population: This indicator is based on US patent data and may favour English-speaking countries.

Ratio of publications in the 19 most industry-relevant scientific disciplines per million population: between 1980 and 1995.

Business enterprise R&D (BERD) as a percentage of GDP: between 1996 and 1999. This covers R&D activities carried out in the business sector regardless of the origin of funding. R&D data are often underestimated, especially in small and medium-sized enterprises (SMEs) and in service industries.

Number of business researchers per 10 000 labour force in 1999: researchers are defined as professionals engaged in the conception and creation of new knowledge, products, processes, methods and systems and are directly involved in the management of projects.

Number of patents in "triadic" patent families per million population: Patent families, as opposed to patents, are a set of patents taken in various countries for protecting a single invention.

Share of firms having introduced at least one new or improved product or process on the market over a given period of time: Is an indicator of the output of innovative activities. This indicator is taken from the Community Innovation Survey (CIS2) managed by Eurostat. It is weighted here by number of employees in order not to underestimate the weight of large firms (unweighted results would give an unduly large weight to the mass of small firms). Data is available only for 21 OECD countries.

A.5 From The World Bank's WDI:

The WB's World Development Indicators are a huge database covering country-level annual indicators for more than 180 countries and aggregates, since 1960. The indicators numbered 25-28, 39-46, 55-57, and 62-65 in Table 4, and 5, 30-33, and 44-65 in Table 10. Their descriptions are the following:

Cost of business start-up procedures (% of GNI per capita): Cost to register a business is normalized by presenting it as a percentage of gross national income (GNI) per capita.

Domestic credit provided by banking sector (% of GDP): Domestic credit provided by the banking sector (monetary authorities, deposit money banks, and other banking institutions) includes all credit to various sectors on a gross basis, except the central government.

Domestic credit to private sector (% of GDP): refers to financial resources provided to the private sector, such as through loans, purchases of nonequity securities, and trade credits and other accounts receivable, that establish a claim for repayment.

Highest marginal tax rate, corporate rate (%): highest rate shown on the schedule of tax rates applied to the taxable income of corporations.

Highest marginal tax rate, individual (on income exceeding, US\$): highest rate shown on the schedule of tax rates applied to the taxable income of individuals.

Interest rate spread (lending rate minus deposit rate): (self-explanatory).

Labor force with primary / secondary / tertiary education (% of total labor force): (self-explanatory).

Management time dealing with officials (% of weekly management time): time dealing with requirements imposed by government regulations (taxes, customs, labor regulations, licensing and registration).

Patent applications, nonresidents / residents: applications filed with a national patent office for exclusive rights for an invention.

Procedures to enforce a contract (number): number of independent actions, mandated by law or courts, that demand interaction between the parties of a contract or between them and the judge or court officer.

Procedures to register property (number): number of procedures required for a businesses to secure rights to property.

Physicians (per 1,000 people): Physicians are defined as graduates of any faculty or school of medicine who are working in the country in any medical field (practice, teaching, research).

Start-up procedures to register a business (number): number of required actions to start a business, including interactions to obtain necessary permits and licenses and to complete all inscriptions, verifications, and notifications to start operations.

Time required to start a business (days): number of calendar days needed to complete the procedures to legally operate a business.

Technicians in R&D (per million people): Technicians in R&D and equivalent staff are people whose main tasks require technical knowledge and experience in engineering, physical and life sciences (technicians), or social sciences and humanities (equivalent staff). They participate in R&D by performing scientific and technical tasks involving the application of concepts and operational methods, normally under the supervision of researchers.

Trademarks, nonresidents / residents: applications for registration of a trademark with a national or regional trademark office.

Taxes on exports (% of tax revenue): (self-explanatory).

Micro, small and medium enterprises (number): (self-explanatory).

Listed domestic companies, total: domestically incorporated companies listed on the country's stock exchanges at the end of the year.

B Brief Overview of Standard PCA

"The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the dataset. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables." (Jolliffe, 2002, chapter 1).

There are many ways for getting the PCs. Here we present the eigen-decomposition of the covariance matrix. Suppose we have a set of high-dimensional datapoints x_i , where x_i is the i^{th} row of the D -dimensional data matrix X . The low dimensional counterpart of x_i is denoted by y_i , where y_i is the i^{th} row of the d -dimensional data matrix Y . Without loss of generality, assume X has zero empirical mean. PCA attempts to find a linear transformation T that maximizes

$$T'cov(X)T$$

under the constraint that $|T| = 1$, and where $cov(X) = E(XX')$ is the covariance matrix of the zero mean data X . This constraint can be enforced by introducing a Lagrange Multiplier λ . Hence, an unconstrained maximization of

$$T'E(XX')T + \lambda(1 + T'T)$$

is performed with respect to T and λ . The first order conditions give us the eigenproblem that PCA solves:

$$E(XX')v = \lambda v$$

The eigenproblem is solved for the d principal eigenvalues λ . The corresponding eigenvectors form the columns of the linear transformation matrix T . The low-dimensional data representations y_i of the data points x_i are computed by mapping them onto the linear basis T , i.e.

$$Y = XT.$$

The weights we are looking for are the square eigenvectors, i.e., for each eigenvalue λ_i , starting with the largest one, there exists an eigenvector T_i that denotes the i^{th} column of T , and the weights corresponding to the i^{th} PC are

$$W_i = (T_i)^2$$

These are the weights we will use to get the standard PCA-based composite indicator:

$$CI_i = X \cdot W_i'$$

If we order the eigenvalues from bigger to smaller, and order the corresponding eigenvectors in T , then choosing the first PC implies choosing the one that accounts for the largest portion of the dataset variance. If we choose the second, we obtain the second largest PC, accounting for the second largest portion of the dataset variance, and so on. The PC is obtained as

$$PC_i = X \cdot T_i'$$

and its variance is the i^{th} eigenvalue, i.e.

$$var[PC_i] = T_i'E(XX')T_i = \lambda_i$$

An additional property is that the sum of the variance of each PC, namely the sum of the eigenvalues, equals the sum of the variance of the elements of X . So if the variance of each element in X is one (as it would be the case if all variables were standardized) then the sum of the eigenvalues should equal the number of elements in X . Furthermore,

$$t_i = \frac{\lambda_i}{\sum_{j=1}^D var(x_j)}$$

is the general expression for the proportion of the overall variance capture by the i^{th} principal component, PC_i .

How many and which components (eigenvectors) should be retained in the analysis without losing too much information is a delicate matter. There is no unified opinion on this²¹ but criteria cover subjective or theoretically-comprehensible ones²², rules-of-thumb²³, graphical analysis²⁴, and other more complex such as bootstrapped eigenvalues and eigenvectors²⁵, cross-validatory criteria²⁶, and intrinsic dimensionality estimation (van der Maaten, 2007).

We use the latter to test out how many eigenvalues adequately represents our datasets, both for the aggregated dataset and for each set of sub-indicators. In general the results vary between 8 and 10 eigenvalues for the overall set of indicators, and between 1.5 and 3 for each dimension, using the Maximum Likelihood intrinsic dimensionality estimator. The relevant principal components for the full set of indicators are in line with the way the dataset was originally constructed, i.e., using 9 categories, with the current first dimension divided in three ones (see Table 3 and 7, sub-dimensions 1.1, 1.2, and 1.3). We also report the classical Scree Plot (see Figure 4) for the full set of indicators, for both periods, for the 10 largest PCs to get an idea on PCs order selection using a well-known method. In spite of these results we will use just the largest PC, in order to obtain a single set of weights.

The estimations of standard PCA have been performed with the Data Reduction Toolbox for Matlab (van der Maaten’s DRToolbox, 2007) which is freely available online. Code was adapted to fit our problem.

C The Multi-block Weighting Approach: Consensus PCA

Consensus PCA was introduced by Wold et al. (1987) as a method for comparing several blocks of descriptor variables measured on the same objects or observations.

Following Westerhuis et al. (1998) and using the notation of the standard PCA, the method can be described in the following way: the data are divided into B blocks X_1, \dots, X_B . A *consensus direction* among all the blocks is sought (i.e. a starting consensus or *super-score* is selected as a column of one of the blocks). This vector is regressed on all blocks X_b , i.e. for each block $X_b, b = 1, \dots, B$, we compute

$$p_b = (X'_b \cdot t_T) / (t'_T t_T)$$

where t_T is the super-score and p_b are the loadings of variables in block X_b . The p_b are normalized to $\|p_b\| = 1$, where $\|\cdot\|$ stands for the Euclidean norm for multidimensional spaces.

From the block variable loadings, block scores t_b for all blocks are calculated, namely

$$t_b = (X_b \cdot p_b) \cdot (m_{X_b})^{-\frac{1}{2}}$$

where $(m_{X_b})^{-\frac{1}{2}}$ is a scaling factor for each block, and m_{X_b} is the number of variables in block X_b . All block scores are combined into a super block T . The super score t_T is then regressed on the super block to give the super weight w_T of each block score to the super score, i.e.

$$\begin{aligned} T &= [t_1 \dots t_B] \\ w_T &= (T' \cdot t_T) / (t'_T t_T) \end{aligned}$$

Then w_T is normalized to $\|w_T\| = 1$ and a new super score is calculated:

$$t_T = T \cdot w_T$$

²¹See Jolliffe (2002, chapter 6). Discussions can also be found in Giri (2004), Jobson (1992), Manly (1994), and in many texts on multivariate analysis and PCA. For an application to social science see Bartholomew et al. (2002, chapter 5).

²²See, for example, Houweling et al. (2003) and Filmer and Pritchett (2001).

²³The most well-known criteria based on the eigenvalues rule-of-thumb (i.e. size of variances of principal components) are the Guttman-Kaiser Criterion (Guttman, 1954; Kaiser, 1960, 1961, Yeomans and Golder, 1982), the broken stick model (Legendre and Legendre, 1983), and the Jolliffe Criterion (Jolliffe, 1972, 1973, 1987). There exists also criteria related to the size of explained variance, cumulative percentage of total variation, and partial correlations.

²⁴Two popular graphical methods are the scree graph, discussed by Cattell (1966), and the log-eigenvalue or LEV diagram (see Farmer, 1971). These methods are very simple but the main drawback is the subjective judgement the analyst has to do to get the number of components to keep.

²⁵See Jackson (1993) and Yu et al. (1998).

²⁶See Wold (1978), and Eastment and Krzanowski (1982). These methods are characterized by their high computational intensity.

The procedure is iterated until t_T converges to a predefined precision (in our example, we use $1e^{-12}$).

The super score is derived using all variables, whereas the block scores are derived using only the variables within the corresponding block. The super weight w_T gives the relative importance of the different blocks X_b for each dimension. After convergence, all blocks are deflated using the super score. Following this algorithm we get two set of weights that are reported in Tables 6 and 12, last two columns.

A brief comment must be done on other method, the Hierarchical PCA or HPCA, that is sometimes confused with CPCA. HPCA was introduced by Wold et al. (1987) as a variant of CPCA, with other normalization rules (namely, t_b and t_T are normalized instead of w_T and p_b). These little differences makes them totally different. In fact, we analyzed the viability of HPCA for our problem of constructing the CEI and we found that many of the highlighted drawbacks of the method apply to us. First, HPCA is sensible to dominant directions: in both algorithms the super score will be the direction most dominant in the consensus block T . However, because in HPCA the block scores are normalized to length one, the algorithm searches for the most dominant direction in these normalized scores. In CPCA the block scores enter T as they are calculated for each block and therefore the super score will just be the direction most dominant in the block scores. Differences between the methods can be expected when a strong direction exists in only a single block, which can be the case of entry or exit rates in the entrepreneurship context. When the directions are spread among the blocks, the methods are expected to give similar results. Clearly, this characteristic violates the Non-Dictatorship Axiom previously stated. Second, HPCA converges to different solutions depending on the starting vector. Moreover, if the starting vector is highly correlated to a dominant direction in one of the blocks, the algorithm cannot escape from this direction and will select it as the direction of the super score t_T . Finally, HPCA has no clear objective function to be maximized, and this prevents to force the algorithm to achieve a "specific solution". All of these drawbacks led us two reject these procedure for estimating our CEI.

The estimations of Consensus PCA have been performed with the Multi-Block Toolbox for Matlab (van den Berg's MBToolbox, 2001) which is freely available online. Code was adapted to fit our problem.

D Probability PCA as a Tool for Dealing with Missing Values

To overcome the problem of missing data we used Probability Principal Components Analysis (PPCA) based on the Expectation-Maximization (EM) algorithm, as developed by Sam Roweis (1998). Roweis stated that PCA can be viewed as a limiting case of a particular class of linear-Gaussian models. The goal of such models is to capture the covariance structure of an observed D -dimensional variable X using fewer than $D(D+1)/2$ free parameters required in a full covariance matrix. Linear-Gaussian models do this by assuming that X was produced as a linear transformation of some d -dimensional latent variable Y plus additive Gaussian noise. Denoting the transformation by the $D \times d$ matrix C , and the (D -dimensional) noise by v (with covariance matrix R) the generative model can be written as

$$Y = CX + v$$

with $x \sim \mathcal{N}(0, I)$ and $v \sim \mathcal{N}(0, R)$. The latent or cause variables Y are assumed to be independent and identically distributed according to a unit variance spherical Gaussian. Since v are also independent and normally distributed (and assumed independent of Y), the model reduces to a single Gaussian model for X which can be written explicitly:

$$X \sim \mathcal{N}(0, CC' + R).$$

Then the proposed EM algorithm has the following steps:

$$\begin{aligned} \mathbf{E}\text{-step} & : Y = (C'C)^{-1}C'X \\ \mathbf{M}\text{-step} & : C^{new} = YX'(XX')^{-1} \end{aligned}$$

where X is a $D \times n$ matrix of all the observed data and Y is a $d \times n$ matrix of the unknown states. The columns of C will span the space of the first d principal components.

To arrange the missing data we have to modify the E-step. Instead of estimating only Y as the value which minimizes the squared distance between the point and its reconstruction we can generalize the E-step in the following way:

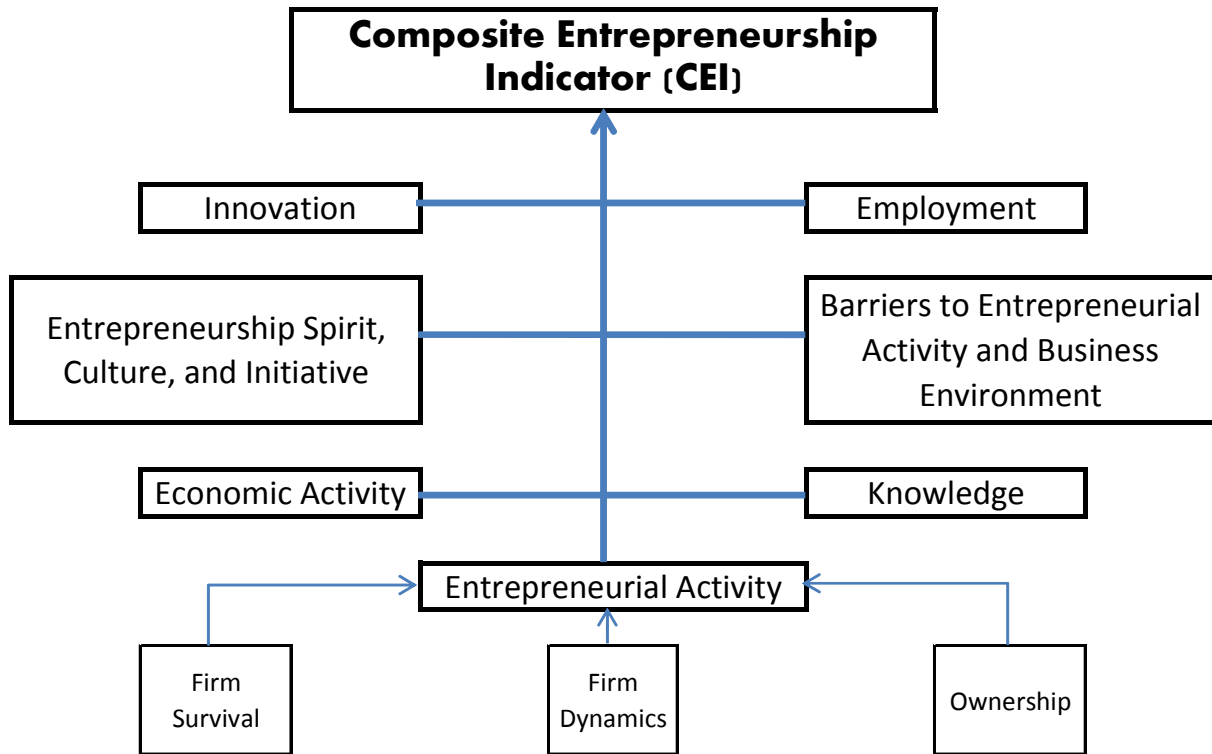
Generalized E-step: For each (possibly incomplete) point X find the unique pair of points Y^* and X^* (such that Y^* lies in the current principal subspace and X^* lies in the subspace defined by the known information about X) which minimize the norm $\|CY^* - X^*\|$. Set the corresponding column of Y to Y^* and the corresponding column of X to X^* .

If X is complete, then $X^* = X$ and Y^* is found exactly as before. If not, then Y^* and X^* are the solution to a least squares problem and can be found by, for example, QR factorization of a particular constraint matrix.

The estimations of missing data based on PPCA have been performed with the `PPCA_MV` procedure implemented in Matlab (Verbeek's `PPCA_MV` Matlab code, 2006) which is freely available online. Code was adapted to fit our problem.

E Tables and Figures

Figure 1.- The Composite Entrepreneurship Indicator's (CEI) Scoreboard



Source: Own elaboration.

Figure 2.- Correlation Matrix for the Whole Set of Indicators
Period 1998-2001

Entrepreneurial Activity									Employment									Economic Activity									Entrepreneurship Spirit, Culture, and Initiative									Barriers to Entrepr. Activity and Business Environment						Knowledge Procurement									Innovation																		
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65					
+																																																																					
	+									+																																																											
		+									+																																																										
			+									+																																																									
				+									+																																																								
					+									+																																																							
						+									+																																																						
							+									+																																																					
								+																																																													

Notes: (+) means correlation above +0.5, and (-) means correlation below -0.5. Shaded areas indicate cross-correlation for each group of variables according to the Scoreboard grouping. Numbers 1 through 65 identify variables as they appear in the Scoreboard.

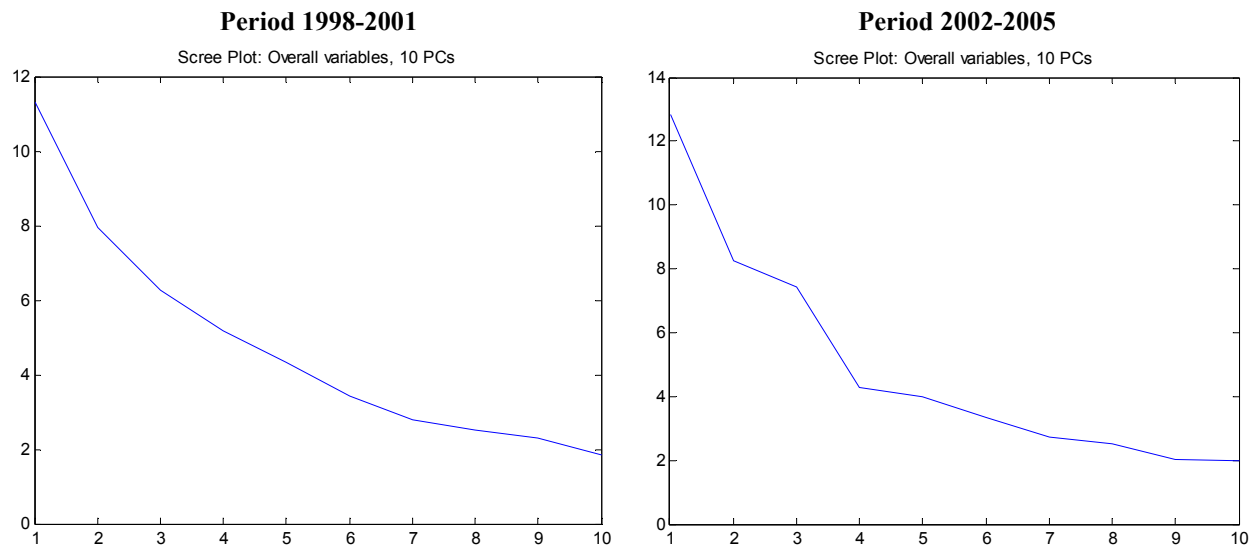
Source: Own calculations. See Appendix A on Data Sources and Variables Description for details.

Figure 3.- Correlation Matrix for the Whole Set of Indicators
Period 2002-2005

Entrepreneurial Activity												Employment											Economic Activity											Entrepreneurship Spirit, Culture, and Initiative											Barriers to Entrepreneurial Activity and Business Environment											Knowledge Procure		Innovation															
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65									
1	+											13											24												34															44															59				62				
2	+	+										13	+											24													34														44														59				62				
3	+	+	+									13	+	+										24														34														44														59				62			
62												13												24													34														44														59				62	+	+	+	
63												13												24													34														44														59				63	+	+	+	
64												13												24													34														44														59				64	+	+	+	
65												13												24													34														44														59				65	+	+	+	

Notes: (+) means correlation above +0.5, and (-) means correlation below -0.5. Shaded areas indicate cross-correlation for each group of variables according to the Scoreboard grouping. Numbers 1 through 65 identify variables as they appear in the Scoreboard.
Source: Own calculations. See Appendix A on Data Sources and Variables Description for details.

Figure 4.- Eigenvalues for Overall-Variables Standard-PCA Estimation



Source: Own calculations. See Appendix B for details.

Table 1.- Overall Entrepreneurship Index
Method I – period 1998-2001

Ranking	Country	Country Code	Index	Ranking	Country	Country Code	Index
1	United States	USA	0.9966	36	Ireland	IRL	-0.0709
2	Switzerland	CHE	0.8925	37	Paraguay	PRY	-0.0724
3	Sweden	SWE	0.8065	38	Syrian Arab Republic	SYR	-0.0779
4	Finland	FIN	0.6078	39	Panama	PAN	-0.0804
5	China	CHN	0.5502	40	El Salvador	SLV	-0.0862
6	Japan	JPN	0.4758	41	Peru	PER	-0.0866
7	Australia	AUS	0.4714	42	Colombia	COL	-0.0896
8	Germany	DEU	0.4062	43	Romania	ROM	-0.0932
9	Korea, Rep.	KOR	0.3362	44	Indonesia	IDN	-0.0980
10	Canada	CAN	0.3223	45	Ecuador	ECU	-0.1004
11	Iceland	ISL	0.2283	46	Denmark	DNK	-0.1033
12	Netherlands	NLD	0.2074	47	Nicaragua	NIC	-0.1054
13	United Kingdom	GBR	0.1608	48	Singapore	SGP	-0.1067
14	France	FRA	0.1577	49	Ukraine	UKR	-0.1073
15	New Zealand	NZL	0.1512	50	Dominican Republic	DOM	-0.1180
16	Malaysia	MYS	0.1235	51	Uruguay	URY	-0.1189
17	Brazil	BRA	0.0727	52	Bulgaria	BGR	-0.1230
18	Hong Kong, China	HKG	0.0688	53	Honduras	HND	-0.1350
19	Lebanon	LBN	0.0455	54	Kazakhstan	KAZ	-0.1433
20	Thailand	THA	0.0417	55	Norway	NOR	-0.1443
21	Egypt, Arab Rep.	EGY	0.0256	56	Venezuela, RB	VEN	-0.1497
22	Argentina	ARG	0.0122	57	Guatemala	GTM	-0.1536
23	Jordan	JOR	0.0006	58	Greece	GRC	-0.1564
24	Italy	ITA	-0.0017	59	Russian Federation	RUS	-0.1685
25	Chile	CHL	-0.0119	60	Estonia	EST	-0.1712
26	India	IND	-0.0169	61	Belgium	BEL	-0.1946
27	Kuwait	KWT	-0.0173	62	Luxembourg	LUX	-0.1986
28	Mexico	MEX	-0.0224	63	Portugal	PRT	-0.3746
29	Israel	ISR	-0.0359	64	Spain	ESP	-0.3796
30	Bolivia	BOL	-0.0388	65	Czech Republic	CZE	-0.4665
31	Saudi Arabia	SAU	-0.0444	66	Slovak Republic	SVK	-0.5015
32	United Arab Emirates	ARE	-0.0447	67	Hungary	HUN	-0.5106
33	Costa Rica	CRI	-0.0612	68	Poland	POL	-0.5590
34	South Africa	ZAF	-0.0678	69	Turkey	TUR	-0.6831
35	Austria	AUT	-0.0696				
				Mean	0.0000	Std. Dev.	0.3064

Source: Own estimations. See Appendix for details.

**Table 3.- Multidimensional Entrepreneurship Index
Method IV – Distribution of Weights among Indicators
and Dimensions – Period 1998-2001**

Dimension	Dimension's Weight	Indicators	Indicator's Weight (Intra-area)	Indicator's Weight (overall)
1.- Entrepreneurial Activity	0.48%	1.1.- Firm Dynamics		
		1 Bankruptcy Rate (%)	0.08%	0.00%
		2 Entry Rate (%)	10.96%	0.05%
		3 Exit Rate (%)	0.00%	0.00%
		4 Share Of Bankruptcies In Firm Exits (%)	43.97%	0.21%
		1.2.- Firm survival		
		5 Young Firm Entrepreneurial Activity Index (Index)	14.55%	0.07%
		6 Total Entrepreneurial Activity Index (Index)	30.43%	0.15%
		1.3.- Ownership		
7 Business Ownership Rate (agriculture, hunting, forestry and fishing, rate)	0.00%	0.00%		
8 Business Ownership Rate (private sector excluding agriculture, hunting, forestry and fishing, rate)	0.00%	0.00%		
9 Business Ownership Rate (total private sector, rate)	0.00%	0.00%		
2.- Employment	2.50%	10 Average Size Of Firm Entries (number)	0.04%	0.00%
		11 Share Of Entries In Employment (%)	0.00%	0.00%
		12 Share Of Exits In Employment (%)	0.00%	0.00%
		13 Aver. Number Workers For Fast Growers; Last Year Period (x 1)	12.47%	0.31%
		14 Aver. Number Workers For All Enterpr.; Growth Rate Period (%)	7.68%	0.19%
		15 Aver. Number Workers For All Enterpr.; Last Year Period (x 1)	1.44%	0.04%
		16 Aver. Number Workers For Fast Growers; Growth Rate Period (%)	0.17%	0.00%
		17 Aver. Number Workers For Not Fast Growers; Growth Rate Period (%)	7.62%	0.19%
		18 Self-employment rates: total, as a percentage of total civilian employment	70.57%	1.77%
3.- Economic Activity	33.02%	19 Average Sales For All Enterprises; Growth Rate Period (%)	0.00%	0.00%
		20 Average Sales For All Enterprises; Last Year Period (x € 1000)	0.00%	0.00%
		21 Average Sales For Fast Growers; Growth Rate Period (%)	0.00%	0.00%
		22 Average Sales For Fast Growers; Last Year Period (x € 1000)	0.00%	0.00%
		23 Average Sales For Not Fast Growers; Growth Rate Period (%)	0.00%	0.00%
		24 Average Sales For Not Fast Growers; Last Year Period (x € 1000)	0.00%	0.00%
		25 Listed domestic companies, total	0.00%	0.00%
		26 Micro, small and medium enterprises (number)	100.00%	33.02%
		27 Micro, small and medium enterprises (per 1,000 people)	0.00%	0.00%
28 Taxes on exports (% of tax revenue)	0.00%	0.00%		
4.- Entrepreneurship Spirit, Culture, and Initiative	18.13%	29 Female Total Entrepreneurial Activity Index (Index)	0.63%	0.11%
		30 Necessity Entrepreneurial Activity Index (Index)	0.55%	0.10%
		31 Opportunity Entrepreneurial Activity Index (Index)	0.11%	0.02%
		32 Potential Entrepreneur Index (Index)	0.00%	0.00%
		33 High-skilled self-employment rates	0.00%	0.00%
		34 Fear Of Failure Index (Index)	0.00%	0.00%
		35 Know Entrepreneur Index (Index)	0.00%	0.00%
		36 Latent entrepreneurship	98.69%	17.89%
		37 Informal Investors Index (Index)	0.00%	0.00%
38 Nascent Entrepreneurial Activity Index (Index)	0.02%	0.00%		
5.- Barriers to Entrepreneurial Activity and Business Environment	14.79%	39 Domestic credit provided by banking sector (% of GDP)	51.90%	7.68%
		40 Domestic credit to private sector (% of GDP)	36.31%	5.37%
		41 Highest marginal tax rate, corporate rate (%)	0.07%	0.01%
		42 Highest marginal tax rate, individual rate (%)	0.78%	0.12%
		43 Interest rate spread (lending rate minus deposit rate)	0.14%	0.02%
		44 Labor force with primary education (% of total)	5.08%	0.75%
		45 Labor force with secondary education (% of total)	4.78%	0.71%
46 Labor force with tertiary education (% of total)	0.93%	0.14%		
6.- Knowledge Procurement	17.53%	47 R&D performed by the non-business sector as a percentage of GDP	0.00%	0.00%
		48 Non-business researchers per10 000 labour force	0.00%	0.00%
		49 Basic research as a percentage of GDP	0.00%	0.00%
		50 PhD graduation rates in science, engineering and health	0.00%	0.00%
		51 Scientific and technical articles per million population	0.98%	0.17%
		52 Business-financed R&D performed by government or higher education as a percentage of GDP	0.00%	0.00%
		53 Scientific papers cited in US-issued patents	0.00%	0.00%
		54 Publications in the 19 most industry-relevant scientific disciplines per million population	0.00%	0.00%
		55 Gross domestic expenditure on R&D (% of GDP)	0.00%	0.00%
56 Physicians (per 1,000 people)	0.00%	0.00%		
57 Technicians in R&D (per million people)	99.01%	17.36%		
7.- Innovation	13.55%	58 BERD as a percentage of GDP	0.00%	0.00%
		59 Business researchers per10 000 labour force	0.00%	0.00%
		60 Number of patents in "triadic" patent families per million population	0.00%	0.00%
		61 Share of firms with new or technologically improved products or processes	0.00%	0.00%
		62 Patent applications, nonresidents	7.51%	1.02%
		63 Patent applications, residents	31.16%	4.22%
		64 Trademarks, nonresidents	0.83%	0.11%
65 Trademarks, residents	60.49%	8.20%		

Source: Own estimations. See Appendix for details.

Table 5.- Overall Entrepreneurship Index
Method I – period 2002-2005

Ranking	Country	Country Code	Index	Ranking	Country	Country Code	Index
1	Indonesia	IDN	0.9061	36	Panama	PAN	-0.0325
2	United States	USA	0.7928	37	Ukraine	UKR	-0.0364
3	China	CHN	0.5728	38	Kuwait	KWT	-0.0430
4	Finland	FIN	0.4601	39	Kazakhstan	KAZ	-0.0491
5	Korea, Rep.	KOR	0.4162	40	Guatemala	GTM	-0.0524
6	Mexico	MEX	0.2943	41	Spain	ESP	-0.0528
7	Brazil	BRA	0.2800	42	Thailand	THA	-0.0534
8	United Kingdom	GBR	0.2322	43	Japan	JPN	-0.0554
9	India	IND	0.2014	44	Jordan	JOR	-0.0601
10	Argentina	ARG	0.1387	45	Turkey	TUR	-0.0679
11	Greece	GRC	0.1031	46	Czech Republic	CZE	-0.0720
12	New Zealand	NZL	0.0864	47	Luxembourg	LUX	-0.0728
13	Australia	AUS	0.0576	48	Costa Rica	CRI	-0.0768
14	Venezuela, RB	VEN	0.0572	49	Uruguay	URY	-0.0771
15	Israel	ISR	0.0534	50	Chile	CHL	-0.0957
16	Russian Federation	RUS	0.0244	51	Bulgaria	BGR	-0.0980
17	Dominican Republic	DOM	0.0171	52	Austria	AUT	-0.1025
18	Bolivia	BOL	0.0164	53	Hong Kong, China	HKG	-0.1056
19	Canada	CAN	0.0084	54	Malaysia	MYS	-0.1083
20	Switzerland	CHE	0.0064	55	Hungary	HUN	-0.1121
21	Egypt, Arab Rep.	EGY	0.0056	56	Colombia	COL	-0.1180
22	Syrian Arab Republic	SYR	0.0045	57	South Africa	ZAF	-0.1282
23	El Salvador	SLV	0.0043	58	Romania	ROM	-0.1356
24	Lebanon	LBN	0.0026	59	Estonia	EST	-0.1359
25	United Arab Emirates	ARE	0.0007	60	Norway	NOR	-0.1456
26	Paraguay	PRY	-0.0055	61	Poland	POL	-0.1468
27	Ireland	IRL	-0.0157	62	France	FRA	-0.1595
28	Iceland	ISL	-0.0202	63	Slovak Republic	SVK	-0.1662
29	Saudi Arabia	SAU	-0.0203	64	Singapore	SGP	-0.2136
30	Ecuador	ECU	-0.0234	65	Italy	ITA	-0.2558
31	Sweden	SWE	-0.0234	66	Belgium	BEL	-0.2705
32	Nicaragua	NIC	-0.0268	67	Denmark	DNK	-0.3477
33	Peru	PER	-0.0296	68	Netherlands	NLD	-0.3921
34	Portugal	PRT	-0.0304	69	Germany	DEU	-0.4761
35	Honduras	HND	-0.0319	Mean	0.0000	Std. Dev.	0.2248

Source: Own estimations. See Appendix for details.

Table 7.- Multidimensional Entrepreneurship Index
Method IV – Distribution of Weights among Indicators
and Dimensions – Period 2002-2005

Dimension	Dimension's Weight	Indicators	Indicator's Weight (Intra-dim)	Indicator's Weight (overall)	
1.- Entrepreneurial Activity	6.01%	1.1.- Firm Dynamics			
		1	Bankruptcy Rate (%)	5.21%	0.31%
		2	Entry Rate (%)	20.95%	1.26%
		3	Exit Rate (%)	1.03%	0.06%
		4	Share Of Bankruptcies In Firm Exits (%)	9.82%	0.59%
		5	New businesses registered (number)	11.11%	0.67%
		1.2.- Firm survival			
		6	Young Firm Entrepreneurial Activity Index (Index)	2.18%	0.13%
		7	Total Entrepreneurial Activity Index (Index)	0.38%	0.02%
		8	Established Businesses Activity Index (Index)	20.43%	1.23%
		1.3.- Ownership			
		9	Business Ownership Rate (agriculture, hunting, forestry and fishing, rate)	18.27%	1.10%
10	Business Ownership Rate (private sector excluding agriculture, hunting, forestry and fishing, rate)	10.43%	0.63%		
11	Business Ownership Rate (total private sector, rate)	0.12%	0.01%		
12	Non-agricultural business ownership rate	0.07%	0.00%		
2.- Employment	22.01%	13	Average Size Of Firm Entries (number)	9.22%	2.03%
		14	Share Of Entries In Employment (%)	1.95%	0.43%
		15	Share Of Exits In Employment (%)	0.50%	0.11%
		16	Aver. Number Workers For Fast Growers; Last Year Period (x 1)	17.87%	3.93%
		17	Aver. Number Workers For All Enterpr.; Growth Rate Period (%)	18.13%	3.99%
		18	Aver. Number Workers For All Enterpr.; Last Year Period (x 1)	13.44%	2.96%
		19	Aver. Number Workers For Fast Growers; Growth Rate Period (%)	0.81%	0.18%
		20	Aver. Number Workers For Not Fast Growers; Growth Rate Period (%)	17.53%	3.86%
		21	Aver. Number Workers For Not Fast Growers; Last Year Period (x 1)	11.82%	2.60%
		22	Non-agricultural self-employment rate	3.01%	0.66%
		23	Self-employment rates: total, as a percentage of total civilian employment	5.73%	1.26%
3.- Economic Activity	13.21%	24	Average Sales For All Enterprises; Growth Rate Period (%)	5.64%	0.74%
		25	Average Sales For All Enterprises; Last Year Period (x € 1000)	6.83%	0.90%
		26	Average Sales For Fast Growers; Growth Rate Period (%)	13.80%	1.82%
		27	Average Sales For Fast Growers; Last Year Period (x € 1000)	2.13%	0.28%
		28	Average Sales For Not Fast Growers; Growth Rate Period (%)	21.82%	2.88%
		29	Average Sales For Not Fast Growers; Last Year Period (x € 1000)	6.08%	0.80%
		30	Listed domestic companies, total	5.70%	0.75%
		31	Micro, small and medium enterprises (number)	29.90%	3.95%
		32	Micro, small and medium enterprises (per 1,000 people)	7.34%	0.97%
		33	Taxes on exports (% of tax revenue)	0.75%	0.10%
4.- Entrepreneurship Spirit, Culture, and Initiative	12.42%	34	Female Total Entrepreneurial Activity Index (Index)	7.71%	0.96%
		35	Necessity Entrepreneurial Activity Index (Index)	17.72%	2.20%
		36	Opportunity Entrepreneurial Activity Index (Index)	1.16%	0.14%
		37	Potential Entrepreneur Index (Index)	4.11%	0.51%
		38	Fear Of Failure Index (Index)	2.33%	0.29%
		39	Know Entrepreneur Index (Index)	19.54%	2.43%
		40	Innovative entrepreneurship	20.24%	2.51%
		41	Future Entrepreneur Index (Index)	10.44%	1.30%
		42	Informal Investors Index (Index)	15.84%	1.97%
		43	Nascent Entrepreneurial Activity Index (Index)	0.91%	0.11%
5.- Barriers to Entrepreneurial Activity and Business Environment	3.05%	44	Cost of business start-up procedures (% of GNI per capita)	10.64%	0.32%
		45	Domestic credit provided by banking sector (% of GDP)	2.25%	0.07%
		46	Domestic credit to private sector (% of GDP)	0.50%	0.02%
		47	Ease of doing business index (1=most business-friendly regulations)	9.30%	0.28%
		48	Highest marginal tax rate, corporate rate (%)	0.77%	0.02%
		49	Highest marginal tax rate, individual rate (%)	0.75%	0.02%
		50	Interest rate spread (lending rate minus deposit rate)	0.30%	0.01%
		51	Labor force with primary education (% of total)	4.17%	0.13%
		52	Labor force with secondary education (% of total)	8.50%	0.26%
		53	Labor force with tertiary education (% of total)	27.23%	0.83%
		54	Management time dealing with officials (% of management time)	2.12%	0.06%
		55	Procedures to enforce a contract (number)	3.01%	0.09%
		56	Procedures to register property (number)	1.14%	0.03%
		57	Start-up procedures to register a business (number)	8.21%	0.25%
		58	Time required to start a business (days)	21.11%	0.64%
6.- Knowledge Procurement	15.86%	59	Gross domestic expenditure on R&D (% of GDP)	24.54%	3.89%
		60	Physicians (per 1,000 people)	40.39%	6.41%
		61	Technicians in R&D (per million people)	35.08%	5.56%
7.- Innovation	27.44%	62	Patent applications, nonresidents	16.19%	4.44%
		63	Patent applications, residents	9.01%	2.47%
		64	Trademarks, nonresidents	36.63%	10.05%
		65	Trademarks, residents	38.17%	10.47%

Source: Own estimations. See Appendix for details.

