Economic development strategies and methods for identifying leading industries and industrial clusters

Thomas J. Webster

Lubin School of Business,
Pace University,
One Pace Plaza,
New York, NY 10038, USA
E-mail: twebster@pace.edu

Abstract: This paper reviews several popular economic development strategies and discusses the practical problem of identifying leading industries and innovation clusters for government regulatory and financial support. Although there is no substitute for in-depth industry-by-industry analysis, several statistical techniques are available that can assist in the identification process, including principal component analysis, *k*-means clustering, hierarchical clustering, medoid partitioning and fuzzy clustering. The Republic of Indonesia is used as a case study to illustrate the strengths and weakness of each of these procedures.

Keywords: economic development; export-led growth; fuzzy clustering; hierarchical clustering; Indonesia; innovation clusters; *k*-means clustering; leading industries; medoid partitioning; principal components.

Reference to this paper should be made as follows: Webster, T.J. (2013) 'Economic development strategies and methods for identifying leading industries and industrial clusters', *Int. J. Economics and Business Research*, Vol. 5, No. 1, pp.55–74.

Biographical notes: Thomas J. Webster is a Professor of Economics in the Lubin School of Business of Pace University. He previously held positions as an International Economist with the US Government and two money centre commercial banks. His research interests include economic development, efficient markets, game theory, and international trade and finance. He received his Undergraduate degree from the School of International Service of American University in Washington, DC and holds a PhD in Economics from the City University of New York.

1 Introduction

Among the most daunting of challenges confronting development policy-makers is selecting the most appropriate economic growth strategy. As a general rule, successful economic growth strategies exploit a country's comparative advantages and industrial synergies within a stable social, political and economic environment. In many developing countries, however, the prerequisites for sustained economic growth are absent. Social and political unrest, lack of personal security, suffocating government regulations and market intervention, crumbling infrastructure, underdeveloped financial institutions to mobilise domestic savings and an arbitrary legal system that fails to protect private property rights prevent economic take-off and rising living standards.

A key to successful economic development is the need to diversify a country's industrial base by targeting leading individual industries, and groups of industries, to serve as engines of economic growth. Unfortunately, identifying these engines of economic growth can be difficult, especially in economies with underdeveloped system of national income accounting. Even when information on the sources of gross domestic product (GDP) and the decomposition of the labour force are highly aggregated, quantitative methods for data manipulation and analysis can be invaluable exploratory tools in the identification process. The results of these techniques, however, should not be taken at face value, but supplemented with an in-depth assessment of the economic, social, political and legal factors to ensure that government support is efficiently targeted.

The topics discussed in this paper are presented as follows. Section 2 reviews several popular development strategies and discusses the practical problem of targeting particular sectors of the economy given a country's national priorities, industrial infrastructure, and science and technology (S&T) base. Sections 3 and 4 survey several exploratory quantitative techniques that may be used to identify leading industries and industrial 'clusters' for government regulatory and financial support. Data on GDP by industrial origin for the Republic of Indonesia will be analysed to illustrate the strengths and limitations of these techniques. Section 5 concludes with a brief discussion of the problems of identifying industries for government support when data on national income, production and employment either non-existent, anecdotal or both.

2 Brief survey of selected economic development strategies

Several theories seeking to explain the sources of sustained economic growth emerged in the latter half of the twentieth century. The most popular development strategies to emerge from these theories include export-led growth, industrial policy and innovation cluster policy. Although each of these strategies emphasises the importance of a country's comparative advantages, there is an ongoing debate over which sectors of an economy should be targeted for preferential treatment and the degree to which government should be directly involved in the marketplace.

2.1 Export-led growth

Export-led growth is a development strategy involving government support for domestic industries that enjoy a comparative advantage in global markets. In theory, government support enables export-oriented domestic industries to exploit economies of scale, which reduce transactions costs and increase international competitiveness. International competition forces domestic producers to adopt the most efficient production and marketing techniques, which spur investment in human capital, promote dynamic innovation and technological diffusion, stimulate domestic production, create employment opportunities and promote price stability (Felipe, 2003).

Export-led growth produces a virtuous cycle of trade, productivity, investment and economic growth, which results in even more trade, etc. Specialisation and trade raises *per capita* productivity and income, and increases domestically sourced savings to finance domestic capital formation. Greater *per capita* income is accompanied by better nutrition, healthcare and expanded educational and training opportunities, which increase worker productivity, lower production costs and improve a country's international competitiveness. Export-led growth also promotes the growth of downstream manufacturing, which enables low-income countries to move up the development ladder. Greater participation by a growing middle class in the political process promotes social and political stability, which attracts foreign direct and portfolio investment to supplement domestic sources of investment financing.

Although the Asian economic 'miracle' of the late-1970s through mid-1990s appeared to validate the wisdom of export-led growth strategies, several economists have argued that this approach to economic development can damage a country's long-term growth prospects. Taylor (1993) emphasised the importance of foreign-exchange earnings to finance capital formation and the importation of capital equipment, essential raw materials (particularly oil) and food. Many low-income countries experienced a savings-investment and foreign-exchange gaps that inhibit economic take-off. The resulting overreliance on foreign loans to finance capital expenditures exposed these countries to currency risks, which resulted in price instability and social unrest. Paley (2002) noted that export-led growth can also inhibit the development of the domestic market, leading to a 'race to the bottom' as firms lower wages and pollute the environment to keep production costs low to remain internationally competitive.

Export-led economic growth strategies have also been criticised for their indiscriminant support of export-oriented industries. Shifting too many resources into export-producing industries can result in 'immiserising' growth (Bhagwati, 1958), which can lead to financial crises as the public sector struggles to service a burgeoning foreign debt. Downturn in export demand often causes developing countries to compensate by intensifying investment in export-producing industries, resulting in overcapacity and deteriorating terms of trade. This is what happened in the Dominical Republic and the Caribbean, when governments targeted labour-intensive textile production (Kaplinsky, 1993) and in the East Asia following the downturn in the global semiconductor market. Erturk (2002) and Blecker (2002, 2003) argued that over-investment was the root cause of the East Asian financial crises of the 1990s. According to Blecker (2003), export-led development strategies inhibit diversification of a country's industrial base, which is necessary for self-sustaining economic growth.

2.2 Industrial policy

In part, industrial policy was a response to many of the criticisms that were levelled at export-led growth strategies. Industrial policy refers to government programmes that support selected domestic industries to achieve a national policy objective, including protecting domestic industries that play a critical role in the national defence, supporting industries with high-value-added per worker,[1] ensuring the survival of critical linkage industries,[2] protecting industries that have been targeted by foreign governments,[3] encouraging the growth of export-oriented industries with high growth rate potential, etc. The policy levers available to government to direct the flow of investment and productive resources to targeted industries include production, export, and research and development subsidies, tax preferences, low interest rate loans, preferential exchange rates, tariff protection and countervailing duties.

Among most compelling reasons for an industrial policy is the existence production synergies, which exist when investments made by one firm generate benefits for other firms in the same industry or firms in other industries engaged in related activities. In the absence of such assistance, investment in research and development tends to be sub-optimal because individual firms cannot capture the full benefits of these expenditures. Industrial policy is most effective when emerging industries are integrated into an ever-widening nexus of interconnections that are characteristic of a modern economy.

Industrial policy also has its detractors. Economists have argued that markets do a better job at efficiently allocating resources than government bureaucrats. Moreover, politicians and bureaucrats do not always act in the nation's best interest, especially when micromanagement by government is involved. Favouritism, nepotism, corruption and influence peddling are incompatible with economic efficiency and maximising social welfare. Industrial policy frequently confers political power on special-interest groups that cannot be justified in economic terms.

To be effective, industrial policy should cultivate a social, political, economic and financial environment that is conducive to free enterprise and encourages more efficient domestic production by investing in roads, airports, telecommunications networks and pre-school to university education. It should encourage domestic and foreign investment by establishing a legal framework that guarantees private property rights, and pursue macroeconomic policies that encourage domestic savings and investment. Most importantly, the government should allow the market to function with minimal interference.

Identifying leading industries, however, is a difficult task since it is not possible to assert that sectors that have grown rapidly in the past will continue to do so in the future. Nor is it always possible to transplant successful industries from other countries.[4] Misdirected resources introduce production inefficiencies and distortions that undermine a country's comparative advantages. Moreover, industrial policy has been criticised for its high social welfare cost. Industrial policy typically results in higher taxes and product prices that benefit domestic producers at the expense of consumers.

2.3 Innovation clusters

The production of final goods and services involves several stages that link upstream raw materials to final distribution and sale – the so-called value chain. Industrial policies that target specific industries, but ignore intra- and inter-industry linkages to lower transactions costs, often produce non-optimal results. For this reason, governments should direct development support to groups of synergistic industries. These clusters may be vertical (industries involving different links in the value chain) or horizontal (firms that constitute the same link in the value chain). This is the central idea underlying the idea of innovation clusters, which was first proposed by Porter (1990, 1996, 1998).

According to Porter, innovation drives productivity, job creation and economic growth. Innovation is not limited to a single firm but requires the participation of interrelated firms to exploit new knowledge and technology. Innovation clusters enhance the value of inter-firm networking and collaboration; lower per unit costs; attract investment in new technology, research and development; foster new capital formation; upgrade managerial and labour skills; and raise living standards. Innovation clusters exploit existing comparative advantages and create new comparative advantages, which are essential for international competitiveness.

The assertion that ‘all clusters matter’, however, is not a development strategy. Innovation at the upper end of the technology spectrum requires a critical mass of interrelated firms and a sufficiently large market to exploit economies of scale and scope (institutional infrastructure); a well-developed S&T base (S&T infrastructure); a culture that is conducive to risk-taking entrepreneurship; and strong forward and backward linkages connecting producers and consumers. Since these prerequisites are satisfied in varying degrees in low- and middle-income countries, innovation policy must be tailored to a country’s stage of economic development (Aubert, 2004).

Low-income countries often lack both the institutional and the S&T infrastructure to support high-technology innovation clusters. When this is the case, innovation policy should focus on the country’s basic technological capability as a means of elevating social welfare, education and training, and agriculture. In contrast, middle-income countries with a basic institutional infrastructure, but a rudimentary S&T base, should focus on emerging technologies such as information systems, data processing, network integration, etc. Upper middle-income countries can move up the development ladder by encouraging investment in its S&T capabilities and emphasising higher-value-added production by investing in human capital, adopting new technologies and promoting entrepreneurship to drive innovation and creativity.

The choice of innovation clusters depends on the government’s development objectives. While export-oriented clusters have become the central organising principal for many developing economies (Bergman, 1998; Bergman and Feser, 1997), several Organisation for Economic Cooperation and Development countries have targeted innovation clusters to reshape regional and national development policies (Roelandt and Hertog, 1998). The US government has embraced the idea of innovation clusters as a prescription for economic development in Africa and war-torn Iraq (USAID, 2006).

As with all development strategies, success requires a stable social, political and cultural environment within which individual initiative, entrepreneurship and unfettered market are allowed to flourish. Uncertainty about private property rights and personal safety can undermine any well-conceived and well-intentioned economic development programme as easily as outright military conflict, civil disorder, supply disruptions and intrusive government interference. Producer and consumer confidence, and a market mechanism that efficiently allocates scarce resources are critical for success.

3 Quantitative methods for identifying leading industries

Several ranking criteria have been suggested for identifying leading industries. The most popular are those with high growth rate potential. Although picking ‘winners’ is a difficult task under the best of circumstances, a popular rule-of-thumb approach is to assume that industries that have grown rapidly in the past will continue to do so in the future. Unfortunately, the checkered track record of these development strategies has demonstrated that this is as much an art as science. Although there is no substitute for an in-depth industry-by-industry analysis, several exploratory statistical techniques may be useful in the identification process.

3.1 Input–output analysis

Input–output analysis has been used extensively to quantify the interrelationships between different sectors of a macroeconomic system. Input–output analysis provides a detailed mathematical description of how materials flow between industries and where additional value can be created.[5] A leading industry must have significant feedback effects with other sectors to serve as an engine of economic growth. The main advantage of input–output analysis is that these interdependencies can be quantified.

A useful by-product of input–output analysis is partial and gross (Keynesian) multipliers, which may be used to define a leading industry. Spending multipliers provide quantitative measures of the impact on the economy from a change in aggregate demand. An incremental change in 'autonomous' spending results in an initial increase in national income, which generates successive rounds of numerically smaller increments of spending and income until a macroeconomic equilibrium has been achieved. Partial multipliers quantify the first-round spending effects and Keynesian multipliers provide a measure of the aggregate change in income and output.

Although the mathematics of input–output analysis is relatively straightforward, the massive data requirements render this approach for identifying leading industries in most low- and middle-income countries impractical. Fortunately, there are other quantitative techniques with less demanding data requirements that are available to assist in the identification process. One of the most promising of these is principal component analysis (PCA).

3.2 Principal component analysis

PCA is a variable-reduction technique that untangles complex patterns in multivariate data suffering from severe multicollinearity.[6] PCA generates vectors of uncorrelated parameter estimates called principal components. There are as many principal components as there are explanatory variables. The first principal component provides the best explanation of the variation in the original variables. The second principal component explains the largest remaining variation, etc. Each successive principal component accounts for less and less variance. What remains is noise or 'scree'. The collection of all principal components accounts for the total variation of the original variables.

The heuristic value of PCA makes it a useful exploratory tool in the identification process. Unlike regression analysis, PCA does not attempt to explain variations in the value of a predetermined dependent variable on the basis of a vector of explanatory variables. Instead, variations in the explanatory variables explain the behaviour of a latent (unidentified) dependent variable (Lattin et al., 2002). The critical step in PCA is the choice of explanatory variables. The revealed decomposition may then be linked to theory through empirical observation and testing.

To illustrate the usefulness of PCA for identifying leading industries, consider the Republic of Indonesia. Indonesia is typical of many developing countries for which data on GDP by industrial origin and the composition and distribution of the labour force is highly aggregated, albeit somewhat more detailed than for many low-income economies. PCA was applied to quarterly data of inflation-adjusted GDP by industrial sector for the period Q1/2000–Q1/2011 published by the Central Bureau of Statistics. Table 1 summarises the 29 industrial sectors analysed in this study.

Table 1 Indonesia's industrial sectors

No.	*Sector*
1	Food crops
2	Estate crops
3	Livestock and products
4	Forestry
5	Fishery
6	Crude petroleum and natural gas
7	Non-oil and gas mining
8	Quarrying
9	Petroleum and refinery
10	Liquefied natural gas
11	Food, beverage and tobacco
12	Textile, leather products and footwear
13	Wood and wood products
14	Paper and printing products
15	Fertilisers, chemical and rubber products
16	Cement and non-metallic quarrying products
17	Iron and steel basic metal
18	Transportation equipment, machinery and apparatus
19	Other manufacturing products
20	Electricity, gas and water supply
21	Construction
22	Wholesale and retail trade
23	Hotels and restaurants
24	Transport
25	Communication
26	Bank and non-bank financial services
27	Building rental and business services
28	Government services
29	Private services

Table 2 reports the first five principal components and the weights (factor loadings) for each explanatory variable. The relative importance of each industrial sector may be gleaned from the absolute values of the factor loadings. The reported principal components are arrayed in descending order according to their ability to explain variations of the latent variable. For example, the first principal component (Prin1) explains almost 80% of the variation in the latent variable. The first five principal components explain more than 96% of the total variation.

Table 2 Principal components of Indonesia's industrial sectors (Q1/2000–Q1/2011)

No.	*Prin1 (79.8%)*	*Prin2 (6.4%)*	*Prin3 (5.9%)*	*Prin4 (2.2%)*	*Prin5 (2.0%)*
1	–0.1641[a]	–0.0271	0.2508	0.4794	–0.1002
2	–0.1403[a]	0.2942[a]	–0.3101[a]	0.3277	–0.3197
3	–0.2050[a]	0.0048	–0.0235	–0.1506	–0.0834
4	0.0817	0.4827[a]	–0.3977[a]	0.0191	–0.3920
5	–0.2060[a]	0.0822	0.0440	–0.0259	0.0092
6	0.1837[a]	0.2688	0.1988	–0.0300	0.0027
7	–0.1932[a]	0.1393	0.0389	–0.0806	–0.2167
8	–0.2050[a]	0.0851	0.0626	–0.1049	0.0189
9	0.1694[a]	0.2062	0.0710	–0.0170	–0.2870
10	0.1950[a]	–0.0634	–0.1195	–0.0023	0.0284
11	–0.0921[a]	0.3524[a]	–0.4410[a]	0.0671	0.6226[a]
12	–0.1359[a]	–0.4200[a]	–0.2575	0.0512	–0.3857
13	0.1431[a]	–0.0418	–0.2314	–0.7132[a]	–0.1087
14	–0.2049[a]	–0.0416	–0.0447	–0.0393	–0.0017
15	–0.2064[a]	–0.0348	–0.0284	–0.0337	0.0957
16	–0.1905[a]	–0.2113	–0.1805	–0.0718	–0.0807
17	0.1455[a]	0.2014	0.4049[a]	–0.1504	–0.1559
18	–0.2066[a]	–0.0316	–0.0074	–0.0600	0.0141
19	–0.1894[a]	–0.1930	–0.1575	0.0576	–0.0072
20	–0.2027[a]	0.1202	0.0728	–0.0764	–0.0085
21	–0.2059[a]	0.0608	0.0467	–0.0841	0.0422
22	–0.2050[a]	0.0628	0.0755	–0.0853	0.0270
23	–0.2059[a]	0.0631	0.0647	–0.0723	0.0029
24	–0.2069[a]	–0.0116	–0.0047	–0.0719	–0.0580
25	–0.1962[a]	0.1658	0.1616	–0.1025	–0.0166
26	–0.2053[a]	0.0388	0.0732	–0.0859	–0.0406
27	–0.2064[a]	0.0498	0.0574	–0.0732	0.0115
28	–0.1947[a]	0.1966	0.1442	–0.0913	–0.0023
29	–0.2062[a]	0.0472	0.0669	–0.0718	–0.0016

Prin, principal component.

[a] Indicates that factor loadings are statistically significant at the 95% confidence level.

Table 3 summarises the ranking of Indonesian industrial sectors from most to least important according to the absolute value of the weighted average of statistically significant factor loadings for the five principal components reported. The industry sector rankings underscore the dominance of downstream production and the relatively less important contributions to real GDP of primary commodities and other labour-intensive sectors.

Table 3 Principal components industry sector rankings (Q1/2000–Q1/2011)

No.	*Sector*	*Rank*[a]
24	Transport	1
18	Transportation equipment, machinery and apparatus	2
15	Fertilisers, chemical and rubber products	3
27	Building rental and business services	3
29	Private services	5
5	Fishery	6
21	Construction	7
23	Hotels and restaurants	7
26	Bank and non-bank financial services	9
3	Livestock and products	10
8	Quarrying	10
22	Wholesale and retail trade	10
14	Paper and printing products	13
20	Electricity, gas and water supply	14
25	Communication	15
10	Liquefied natural gas	16
28	Government services	17
7	Non-oil and gas mining	18
16	Cement and non-metallic quarrying products	19
19	Other manufacturing products	20
6	Crude petroleum and natural gas	21
17	Iron and steel basic metal	22
12	Textile, leather products and footwear	23
9	Petroleum and refinery	24
1	Food crops	25
2	Estate crops	26
13	Wood and wood products	27
11	Food, beverage and tobacco	28
4	Forestry	29

[a]Rankings based on the absolute value of the weighted average of factor loadings in Table 2.

The two leading industrial sectors for the period examined were transport (1) and transportation equipment (2). Four of the top ten sectors involve services including personal services (5), hotels, restaurants, and bank and non-bank financial services (9). The least important sectors in terms of their contributions to GDP involve the production of primary commodities and labour-intensive goods, including non-oil and gas mining (18), cement and non-metallic quarrying products (19), crude petroleum and natural gas (21), iron and steel (22), textiles and related products (23), food crops (25), estate crops (26), wood and wood products (27), food, beverages and tobacco (28) and forestry (29).

The low ranking of the crude petroleum and natural gas is surprising given Indonesia's membership in the Organisation of Petroleum Exporting Countries and the historical importance of this sector as a leading foreign-exchange earner.

The rankings presented in Table 3 suggest that Indonesia has a diversified industrial base, but a relatively insignificant high-technology manufacturing capability, which is typical of many low- to low-middle-income countries. According to Central Intelligence Agency's World Factbook, Indonesia ranked 128 out of 193 countries examined in terms *per capita* purchasing-power-parity GDP in 2010 of $4,200. This compares with $47,200 in the US (9), $14,700 in Malaysia (57), $3,800 in Iraq (131) and $300 in the Democratic Republic of the Congo (193). In terms of an appropriate innovation policy, Indonesia's stage of economic development suggests that the development efforts should focus on emerging technologies and expanding the country's S&T capabilities. To explore this issue further, we will apply cluster analysis to help identify Indonesian industrial clusters.

4 Quantitative methods for identifying industrial clusters

Cluster analysis refers to a broad group of statistical techniques for sorting objects (members) into groups (clusters) according to shared characteristics that interact for mutual advantage.[7] Cluster analysis does a good job at defining membership if the statistical distance of each member from the centre of a cluster is much less than the distance from the centre of other clusters. If the within-cluster distance is similar to between-cluster distances then cluster membership is ambiguous. Members within each cluster should be as similar as possible, while members of different clusters should be as dissimilar as possible.

There are several exploratory procedures that may be used to identify industrial clusters, including *k*-means clustering, hierarchal clustering, medoid partitioning and fuzzy clustering. Each of these procedures will be illustrated by assigning Indonesian industry sectors to clusters according to their contributions to real GDP. As with PCA, this is not the only sorting criterion. Other membership criteria may include research and development expenditures, per-worker value-added, spillover effects, job creation, forward and backward linkages, etc.

4.1 k-Means clustering

As with all clustering techniques, *k*-means clustering determines optimal cluster membership by minimising within-cluster distances and maximising between-cluster differences.[8] *k*-Means clustering is a hard clustering technique, since members are unambiguously assigned to a specific cluster. This can be a problematic when members are assigned to clusters on the basis of a small number of shared characteristics since disparate members may end up in the same cluster.

Table 4 summarises the results of *k*-means clustering of Indonesian industrial sectors for the period Q1/2000–Q1/2011. The strength of cluster membership can be inferred from the value of the distance statistic (d).[9] Small d values indicate that the member statistically close to the centre of the cluster. Thus, we should more confident that the fertilisers, chemical and rubber products sector with a d value of 0.32 belong to cluster C than we would the transport sector, which has a d value of 1.11.[10] Overall confidence in an identified cluster may be inferred from the average *d* for each cluster. Thus, clusters B (0.66), C (0.64) and D (0.68) appear to be better defined than clusters A (1.21) and E (0.84).

As with all sorting techniques, *k*-means clusters should not be taken at face value. Since the maximum number of possible clusters can be quite large, the objective is to find a 'locally' optimal number of clusters. Thus, membership in each cluster should be closely scrutinised for internal consistency. For example, combining the livestock and fishery in cluster B seems appropriate, but including estate crops and liquefied natural gas in that cluster does not. Similarly, including of non-oil and gas mining, fertilisers, chemical and rubber products in cluster C may be justified as upstream industries the value chain. On the other hand, hotels and restaurants, and transport are downstream industries that probably belong to their own cluster, like tourism.

Table 4 *k*-Means clustering of Indonesian industrial sectors (Q1/2000–Q1/2011)

No.	*Cluster*	*Sector*	*Distance (d)*[a]	*Average distance*
1	A	Food crops	1.21 (3.43)	1.21
11		Food, beverage and tobacco	1.21 (3.55)	
2	B	Estate crops	0.71 (3.11)	0.66
3		Livestock and products	0.38 (2.10)	
5		Fishery	0.61 (2.95)	
10		Liquefied natural gas	0.94 (1.75)	
7	C	Non-oil and gas mining	0.44 (2.30)	0.64
12		Textile, leather products and footwear	0.84 (2.14)	
15		Fertilisers, chemical and rubber products	0.32 (2.72)	
23		Hotels and restaurants	0.50 (2.16)	
24		Transport	1.11 (2.10)	
4	D	Forestry	0.34 (2.18)	0.68
8		Quarrying	0.28 (2.41)	
9		Petroleum and refinery	0.89 (1.63)	
13		Wood and wood products	0.68 (1.81)	
14		Paper and printing products	1.00 (1.46)	
16		Cement and non-metallic quarrying products	0.17 (2.55)	
17		Iron and steel basic metal	0.85 (3.30)	
19		Other manufacturing products	1.47 (3.91)	
20		Electricity, gas and water supply	0.47 (2.86)	
26	E	Bank and non-bank financial services	0.56 (2.80)	0.84
27		Building rental and business services	0.86 (3.81)	
28		Government services	1.04 (3.88)	
29		Private services	0.90 (2.42)	

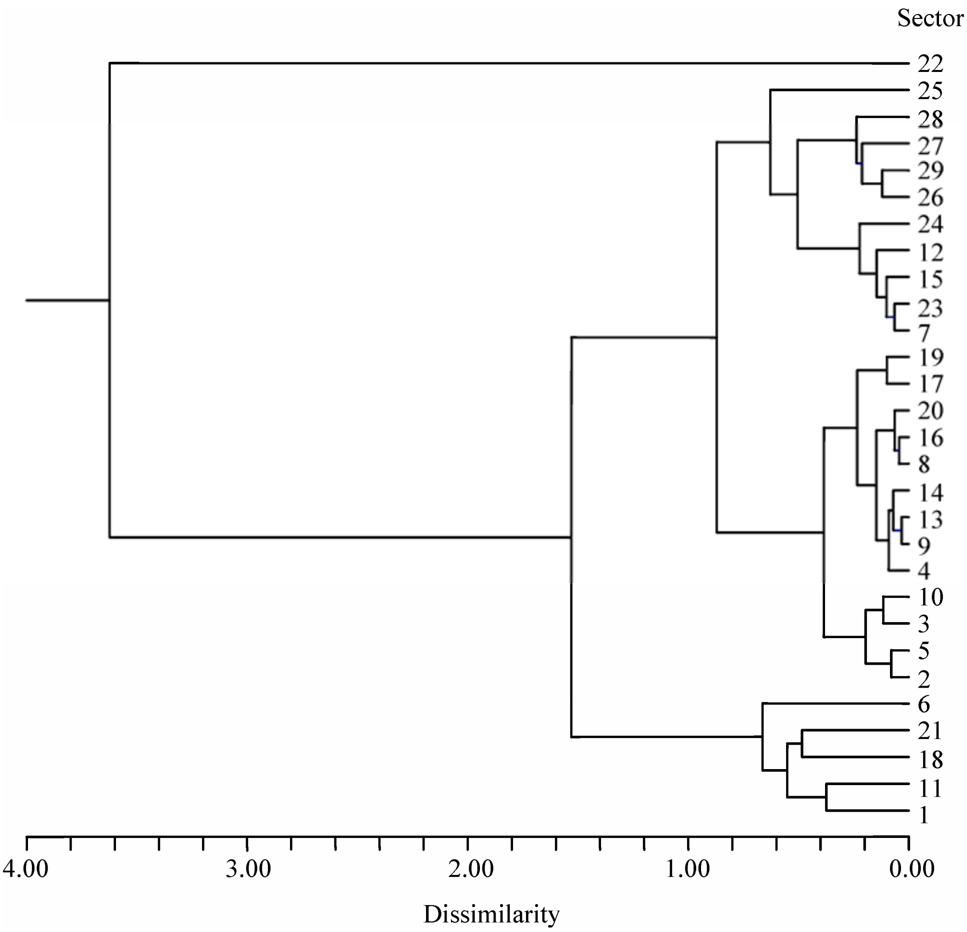
[a] This statistic reports the minimum distance from the centre of a cluster. If the distance from the centre of the cluster is much less than the distance from the centre of the other clusters, the procedure does a good job of clustering. The number in parentheses is the distance to the centre of the next nearest cluster.

4.2 Hierarchical clustering

Hierarchical cluster analysis (Kaufman and Rousseeuw, 1990) is another hard clustering technique that is similar to *k*-means clustering. The distinguishing feature of hierarchical clustering is that it produces a hierarchy of clusters that is displayed in a tree diagram called a dendrogram.

Hierarchical clustering may proceed in two ways. Divisive hierarchical clustering begins by assigning all members to a single, all-inclusive cluster, which is iteratively subdivided into smaller clusters. This process is repeated until an optimal number of clusters are identified. Agglomerative hierarchical clustering proceeds in the opposite direction by first identifying an optimal number of clusters, which are iteratively combined into a single, all-inclusive cluster. Figure 1 depicts the dendrogram for Indonesian industrial sectors. The vertical axis identifies the 29 industry sectors being clustered. The horizontal axis measures the distance between clusters. Division into smaller clusters results in a greater number of branches as we move from left to right.

Figure 1 Dendrogram for Indonesia's industrial sectors (Q1/2000–Q1/2011)



With the other clustering methodologies discussed in this section, the analyst first specifies the number of clusters to be identified, after which within-cluster and between-cluster distances are calculated. With hierarchical clustering, distance values are specified to yield a desired number of clusters. Distance values may be inferred from an examination of the dendrogram. Increasing the distance cut-off value reduces the number of clusters. By setting the distance cut-off value at 0.37 five clusters are identified, which are summarised in Table 5. The membership assignments are broadly consistent with those generated using *k*-means clustering – the exception being that two fewer members (cement and non-metallic quarrying products, and iron and steel basic metal) in cluster D. The strength of cluster memberships based on calculated *d* statistics are also consistent with *k*-means clustering.

Table 5 Hierarchical clustering of Indonesia's industrial sectors (Q1/2000–Q1/2011)

No.	*Cluster*	*Sector*	*Distance (d)*
1	A	Food crops	0.37
11		Food, beverage and tobacco	
2	B	Estate crops	0.20
3		Livestock and products	
5		Fishery	
10		Liquefied natural gas	
7	C	Non-oil and gas mining	0.22
12		Textile, leather products and footwear	
15		Fertilisers, chemical and rubber products	
23		Hotels and restaurants	
24		Transport	
4	D	Forestry	0.15
8		Quarrying	
9		Petroleum and refinery	
13		Wood and wood products	
14		Paper and printing products	
16		Cement and non-metallic quarrying products	
20		Electricity, gas and water supply	
26	E	Bank and non-bank financial services	0.24
27		Building rental and business services	
28		Government services	
29		Private services	

4.3 Medoid partitioning

Medoid partitioning (Kaufman and Rousseeuw, 1990; Späth, 1985) is another hard clustering technique that identifies clusters in terms of representative members called medoids. A medoid is a cluster for which the average dissimilarity with all other members is minimised. Once the number of clusters is specified, members are assigned to its nearest medoid.

The strength of cluster membership may be evaluated using silhouette value (s), which ranges from −1 to 1.[11] A silhouette value close to 1 means that cluster membership is well-defined. A silhouette value close to −1 indicates that there is little evidence for cluster membership. The rule of thumb is that a silhouette value greater than 0.5 is evidence that cluster membership is properly identified.

Table 6 summarises the results of medoid partitioning Indonesia's industrial sectors. Cluster membership is broadly consistent with k-means and hierarchical clustering. Unlike k-means and hierarchical clustering, liquefied natural gas is excluded from cluster B and transport is excluded from cluster C. On the other hand, the iron and steel basic metal sector and manufacturing products sector are included, but the paper and printing products sector excluded.

Table 6 Medoid partitioning of Indonesia's industrial sectors (Q1/2000–Q1/2011)

No.	*Cluster*	*Sector*	*s*	*Averages*
1	A	Food crops	0.71	0.71
11		Food, beverage and tobacco	0.70	
3	B	Livestock and products	0.60	0.56
2		Estate crops	0.51	
5		Fishery	0.58	
7	C	Non-oil and gas mining	0.68	0.64
12		Textile, leather products and footwear	0.53	
15		Fertilisers, chemical and rubber products	0.72	
23		Hotels and restaurants	0.63	
4	D	Forestry	0.66	0.65
8		Quarrying	0.69	
13		Wood and wood products	0.52	
16		Cement and non-metallic quarrying products	0.72	
17		Iron and steel basic metal	0.66	
19		Other manufacturing products	0.58	
20		Electricity, gas and water supply	0.71	
26	E	Bank and non-bank financial services	0.64	0.53
27		Building rental and business services	0.50	
28		Government services	0.49	
29		Private services	0.48	

4.4 Fuzzy clustering

The objective of each of the clustering methodologies discussed is to sort members into distinct clusters. By contrast, fuzzy clustering calculates the probability of membership in each cluster – a procedure known as 'fuzzification' (Kaufman and Rousseeuw, 1990). The sum of these probabilities equals unity. The most likely cluster assignment has the greatest probability of membership. Table 7 summarises the results of fuzzy clustering Indonesia's industrial sectors. Only those clusters with significant silhouette values are reported. Table 7 also summarises the probabilities that the sectors belong to a reported cluster.

Unlike the sorting methods discussed above, which produced five clusters, fuzzy clustering identifies just three industrial clusters. With the exception of liquefied natural gas, which was assigned to cluster D, the sectors belonging to clusters A and B were not assigned. Membership in the remaining clusters is consistent with our earlier results, although cluster D membership is more inclusive.

Table 7 Fuzzy clustering of Indonesia's industrial sectors (Q1/2000–Q1/2011)

				*Probability in cluster**		
No.	*Cluster*	*Sector*	*s*	*C*	*D*	*E*
7	C	Non-oil and gas mining	0.62	*0.59*	0.14	0.14
12		Textile, leather products and footwear	0.59	*0.34*	0.23	0.21
15		Fertilisers, chemical and rubber products	0.52	*0.36*	0.22	0.23
23		Hotels and restaurants	0.64	*0.40*	0.21	0.20
4	D	Forestry	0.78	0.23	*0.38*	0.20
8		Quarrying	0.78	0.24	*0.36*	0.20
9		Petroleum and refinery	0.72	0.23	*0.41*	0.19
10		Liquefied natural gas	0.48	0.25	*0.36*	0.20
13		Wood and wood products	0.75	0.22	*0.42*	0.18
14		Paper and printing products	0.68	0.17	*0.56*	0.14
16		Cement and non-metallic quarrying products	0.79	0.24	*0.37*	0.20
17		Iron and steel basic metal	0.73	0.25	*0.34*	0.21
19		Other manufacturing products	0.67	0.25	*0.33*	0.22
20		Electricity, gas and water supply	0.77	0.24	*0.36*	0.20
26	E	Bank and non-bank financial services	0.77	0.22	0.18	*0.35*
27		Building rental and business services	0.60	0.25	0.20	*0.31*
28		Government services	0.60	0.25	0.20	*0.30*
29		Private services	0.72	0.16	0.14	*0.51*

*Greatest probabilities are italicised for emphasis.

4.5 Summary

Table 8 summarises cluster membership for Indonesia's industrial sectors from each of the partitioning methodologies discussed in this section. Although these techniques produce broadly similar cluster assignments, caution should be exercised when interpreting these results. To begin with, data on Indonesian industrial production is not sufficiently disaggregated to permit identification of intra-industry clusters. Moreover, the assignment of members into clusters is based on shared statistical regularities, which may be coincidental. Thus, caution should be exercised when determining the viability of cluster membership identified by these procedures. In short, clustering should be viewed as an exploratory procedure that should be combined with a detailed industry-by-industry analysis.

The resulting clusters can also be used to assess a country's institutional and S&T infrastructure when formulating an innovation policy. For example, Indonesia appears to have a strong institutional structure, but the absence of a well-developed, high-technology industrial sector suggests that the government should emphasise emerging sectors like information technology and provide incentives to expand its S&T capabilities and diversify its manufacturing sector.

Table 8 Comparison of clustering methodologies

Method	Cluster				
	A	*B*	*C*	*D*	*E*
k-Means	1, 11	2, 3, 5,10	7, 12, 15, 23, 24	4, 8, 9, 13, 14, 16, 17, 19, 20	26, 27, 28, 29
Hierarchical	1, 11	2, 3, 5, 10	7, 12, 15, 23, 24	4, 8, 9, 13,14, 16, 20	26, 29, 28, 29
Medoid	1, 11	2, 3, 5	7, 12, 15, 23	4, 8, 13, 16, 17, 19, 20	26, 27, 28, 29
Fuzzy			7, 15, 23, 24	4, 8, 9, 10, 13, 14, 16, 17, 19, 20	26, 27, 28, 29

5 Conclusions

This study reviewed several popular economic development strategies. An export-led growth strategy emphasises government support of industrial and commercial enterprises that enjoy a comparative advantage in global markets. Related to this is industrial policy, which targets selected industries to achieve a well-defined national policy objective like export-oriented industries with high growth rate potential.

While industrial policy focuses on specific industries, innovation policy targets 'clusters' of industries to exploit positive externalities and spillover effects. Inter-firm networking and collaboration generate benefits that radiate throughout the economy. In the absence of government support, investments by individual firms will be less than optimal. As these synergies develop, emerging industries are integrated into an ever-widening nexus of interconnections that are characteristic of a modern economy.

Successful innovation policy requires a critical mass of interrelated firms and a sufficiently large market to exploit economies of scale and scope. Low-income countries that lack an institutional and S&T infrastructure should focus on basic technology to elevate social welfare, education and training, and agriculture. Middle-income countries with an institutional infrastructure, but rudimentary S&T base, should focus on emerging sectors, like information technology, by investing in human capital, adopting new technologies and promoting entrepreneurship to drive innovation and creativity.

The challenge confronting development policy-makers is to identify the most promising industries and groups of industries to receive government regulatory and financial support. Although there is no substitute for in-depth industry-by-industry analysis, several statistical techniques are available that can assist in the identification process, including input–output and PCA. Unfortunately, the massive data requirements of input–output analysis render this approach impractical for most developing countries. By contrast, the data requirements for PCA are far less demanding.

PCA generates vectors of uncorrelated parameter estimates called principal components. The heuristic value of PCA makes it a useful exploratory tool in the identification process. Unlike regression analysis, PCA does not attempt to explain variations in the value of a dependent variable on the basis of a vector of explanatory variables. Instead, variations in the explanatory variables explain the behaviour of a latent (unidentified) dependent variable. The revealed decomposition may then be linked to theory through empirical observation and testing.

Cluster analysis refers to a broad group of statistical techniques for sorting objects (members) into groups (clusters) with shared characteristics. Revealed clusters may be used to formulate an innovation policy. This paper applied surveyed several clustering techniques, including *k*-means clustering, hierarchical clustering, medoid partitioning and fuzzy clustering.

The Republic of Indonesia was chosen as a case study to illustrate the strengths and weakness of each of the statistical procedures discussed in this paper. Indonesia is typical of many low-income countries for which national income and output data by industrial origin is highly aggregated, albeit somewhat more detailed than for many low-income economies. The analyses were applied to quarterly observations of inflation-adjusted GDP by industrial origin for the period Q1/2000–Q1/2011.

The results of each of the statistical techniques surveyed in this paper should be interpreted with caution. Although PCA is parsimonious in terms of its modest data requirements, the results are sensitive to small changes in the data set. Cluster analysis is a useful exploratory technique for assigning members to clusters based on shared statistical regularities, which may be coincidental. For these reasons, the results of these procedures should not be taken at face value but combined with detailed, industry-by-industry, analyses to provide insights into a country's comparative advantages.

When national income and industrial data are incomplete or unreliable, a more subjective approach to identifying leading industries and clusters may be required. For example, the US government recently undertook to suggest an economic development programme to create employment opportunities in Iraq. Because of the almost complete absence of meaningful national income and industry data, USAID (2006) chose to identify leading industries and innovation clusters using a qualitative approach based on the guidelines suggested by Porter (1990, 1996, 1998). Unfortunately, the recommendations of the USAID study must be taken with a grain of salt.[12]

Although the USAID study provides useful information about the structure of the Iraqi economy and did a good job identifying the economic and social challenges confronting the government, serious internal socio-political problems need to be resolved, including the sectarian violence and an uncertain investment climate. Moreover, Iraq is a low-income country with an adequate institutional structure, but with very basic S&T capabilities. For this reason, an export-led industrial policy that emphasises Iraq's comparative advantages in low-skilled labour, centralised location as a crossroads for international trade and non-oil factor endowments would appear to form the basis of a long-term plan of economic development and job creation.

Acknowledgements

The author would like to thank the anonymous referees of an earlier draft of this paper. Their helpful comments and suggestions were most appreciated.

References

Aldenderfer, M.S. and Blashfield, R.K. (1984) 'Cluster analysis', Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07-044, Newbury Park, CA: Sage Publications.

Armstrong, A.G. and Upton, D.C. (1969) 'A review of input-output applications', *Bulletin of the International Statistical Institute*, Vol. 43, No. 1, pp.113–130.

Aubert, J. (2004) *Promoting Innovation in Developing Countries: A Conceptual Framework*, World Bank Institute, July, Available at: http://info.worldbank.org/etools/docs/library/137729/0-3097AubertPaper%5B1%5D.pdf.

Bergman, E.M. (1998) 'Industrial trade clusters in action: seeing regional economies as a whole', in M. Steiner (Ed.), *Clusters and Regional Specialization: On Geography, Technology, and Networks*, Vol. 8. Pion, London: European Research in Regional Science.

Bergman, E.M. and Feser, E.J. (1997) 'Industrial, regional, or spatial clustering?', OECD Workshop Position Paper on Cluster Analysis and Cluster-Based Policies, Amsterdam 10-11.10.97, Paris: OECD Industrial Cluster Focus Group.

Bhagwati, J. (1958) 'Immiserizing growth: a geometrical note', *Review of Economic Studies*, Vol. 25, No. 3, pp.201–205.

Blecker, R. (2002) 'The balance-of-payments-constrained growth model and the limits to export-led growth', in P. Davidson (Ed.), *A Post Keynesian Perspective on Twenty-First Century Economic Problems*. Northampton, MA: Edward Elgar, pp.69–88.

Blecker, R. (2003) 'The diminishing returns to export-led growth', in W.R. Mead and S.R. Schwenninger (Eds.), *The Bridge to a Global Middle Class: Development, Trade, and International Finance*. Boston, MA: Kluwer, pp.259–298.

Borland, J., Hirschberg, J.G and Lye, J.N. (2001) 'Data reduction of discrete responses: an application of cluster analysis', *Applied Economics Letters*, Vol. 8, No. 3, pp.149–153.

Chatterjee, S. and Price, B. (1977) *Regression Analysis by Example*. New York: John Wiley & Sons.

Dorfman, R., Samuelson, P.A. and Solow, R.M. (1958) *Linear Programming and Economic Analysis*. New York, NY: McGraw Hill.

Erturk, K.A. (2002) 'Overcapacity and the East Asian crisis', *Journal of Post Keynesian Economics*, Vol. 24, No. 2, pp.253–276.

Everitt, B.S., Landau, S. and Leese, M. (2001) *Cluster Analysis* (4th ed.). London, UK: Edward Arnold Publishers, Ltd.

Felipe, J. (2003) 'Is export-led growth passé? Implications for developing Asia', Economics and Research Development Working Paper Series No. 48, Manila, PH: Asian Development Bank.

Gale, D. (1960) *The Theory of Linear Economic Models*. New York, NY: McGraw-Hill.

Hair, J.F., Anderson, R.E., Tathan, R.L. and Black, W.C. (1995) *Multivariate Data Analysis*. Englewood Cliffs, NJ: Prentice-Hall.

Hartigan, J.A. and Wong, M.A. (1979) 'A k-means clustering algorithm', *Applied Statistics*, Vol. 28, No. 1, pp.100–108.

Hotelling, H. (1936) 'Analysis of a complex of statistical variables into principal components', *Journal of Educational Psychology*, Vol. 24, pp.417–441, 498–520.

Kachigan, S.K. (1982) *Multivariate Statistical Analysis*. New York, NY: Radius Press.

Kaplinsky, R. (1993) 'Export processing zones in the Dominican Republic: transforming manufacturers into commodities', *World Development*, Vol. 21, No. 11, pp.1851–1865.

Kaufman, L. and Rousseeuw, P.J. (1990) *Finding Groups in Data: An Introduction to Cluster Analysis*. New York, NY: John Wiley & Sons.

Lattin, J.M., Carrol, J.D. and Green, P.E. (2002) *Analyzing Multivariate Data*. Pacific Grove, CA: Brooks/Cole-Thomson Learning.

Leontief, W. (1970) 'The dynamic inverse', in A.P. Carter and A. Brody (Eds.), *Contributions to Input-Output Analysis*. Amsterdam, NL: North Holland, pp.17–46. Reprinted in Leontief, W. (1986) *Input-Output Analysis*. New York, NY: Oxford University Press, pp.294–320.

Lloyd, S.P. (1957) 'Least squares quantization in PCM', *IEEE Transactions on Information Theory*, Vol. 28, No. 2, pp.129–137, DOI: 10.1109/TIT.1982.1056489, Available at: http://www.cs.toronto.edu/~roweis/csc2515-2006/readings/lloyd57.pdf.

MacQueen, J.B. (1967) 'Some methods for classification and analysis of multivariate observations', *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Berkeley, CA, pp.281–297.

Maddala, G.S. (1992) *Econometrics*. New York, NY: McGraw-Hill.

Malinvaud, E. (1997) *Statistical Methods in Econometrics* (2nd ed.). Amsterdam, NE: North-Holland.

Morishima, M. (1964) *Equilibrium, Stability, and Growth: A Multisectoral Analysis*. London, UK: Oxford University Press.

Paley, T.I. (2002) 'A new development paradigm: domestic demand-led growth', FPIF Discussion Paper, Available at: http://www.fpif.org/papers/development_body.html.

Pearson, K. (1901) 'On lines and planes of closest fit to systems of points in space', *Philosophical Magazine*, Vol. 2, No. 6, pp.559–572.

Porter, M.E. (1990) *The Competitive Advantage of Nations*. New York, NY: Free Press.

Porter, M.E. (1996) 'Competitive advantage, agglomeration economies, and regional policy', *International Regional Science Review*, Vol. 19, Nos. 1–2, pp.85–90.

Porter, M.E. (1998) 'Clusters and the new economics of competition', *Harvard Business Review*, Vol. 76, No. 6, pp.77–90.

Roelandt, T.J.A. and Hertog, P. (1998) 'Cluster analyses and cluster-based policy in OECD countries', in T.J.A. Roelandt and P. Hertog (Eds.), *Cluster Analysis and Cluster-Based Policy: New Perspectives and Rationale in Innovation Policy*. Paris, FR: Organization for Economic Cooperation and Development.

Späth, H. (1985) *Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples*. New York, NY: Halstead Press.

Steinhaus, H. (1956) 'Sur la division des corps matériels en parties', *Bulletin L'Acadmie Polonaise des Science*, Classe III, Vol. IV, No. 12, pp.801–804.

Taylor, L. (1993) 'A three-gap analysis of foreign resource flows and developing country growth', in L. Taylor (Ed.), *The Rocky Road to Reform: Adjustments, Income Distribution and Growth in the Developing World*. Cambridge, MA: MIT Press.

USAID (2006) *Iraq Private Sector Growth and Employment Generation: Iraq Competitive Analysis*. Washington, DC: U.S. Agency for International Development, Available at: http://trade.gov/static/iraq_industrycompetitiveness.pdf.

Notes

[1] Value-added per worker refers to the difference in value between labour used in production and the output produced.

[2] Forward and backward linkage industries, such as steel and shipbuilding or semiconductors and computers, are important because they provide inputs into other industries.

[3] This is a domestic industry that has been targeted by a foreign government for trade promotion and economic development.

[4] For example, the myth behind Japanese industrial policy is that the Ministry of International Trade and Industry (MITI) has an unblemished record of picking 'winners'. In fact, MITI saw no future in automobiles and consumer electronics – industries that Japan came to dominate globally. Sony and Honda flourished despite the absence of government support. In contrast, industries targeted by MITI, such as aluminium, shipbuilding, civilian air transportation and computers, failed to take-off. Japan's economic success had more to do with its stable business environment, highly trained and well-educated labour force, and high savings and investment rates, rather than to its ability to distinguish industrial winners from losers.

[5] See, e.g. Armstrong and Upton (1969), Dorfman et al. (1958), Gale (1960), Leontief (1970) and Morishima (1964).

[6] See, e.g. Chatterjee and Price (1977), Hair et al. (1995), Hotelling (1936), Maddala (1992), Malinvaud (1997) and Pearson (1901).

[7] See, e.g. Aldenderfer and Blashfield (1984), Borland et al. (2001), Everitt et al. (2001), Kachigan (1982), and Kaufman and Rousseeuw (1990).

[8] See, e.g. Hartigan and Wong (1979), Lloyd (1957), MacQueen (1967) and Steinhaus (1956).

[9] The distance between members depends on the nature of the data. For interval variables, the distance between objects is the difference in their values, which is often measured in terms of standard deviation (Euclidean) or average absolute deviation (Manhattan). For a more detailed discussion, see Späth (1985), and Kaufman and Rousseeuw (1990).

[10] There are two possible remedies if cluster membership appears unnatural. Either increase the number of sorting characteristics or split the clusters into two or more sub-clusters according to the information from other sources, analyst expertise or qualitative guidelines.

[11] Kaufman and Rousseeuw' (1990) silhouette values were used to define cluster membership. These silhouette values (s) are constructed by calculating the smallest average dissimilarity (a) of members in cluster A, calculating the smallest average dissimilarity (b) of individual members in cluster B, etc. If $a < b$, $s = 1 - a/b$. If $a > b$, $s = b/a - 1$. If a cluster contains just one member or if $a = b$, $s = 0$. When s is close to 1, cluster membership is well-defined. When s is close to -1, the cluster membership is poorly defined.

[12] The accepted practice when identifying innovation clusters is to examine industry contributions to GDP and employment data, as well as the share of the country's trade in the world economy. In the case of Iraq, however, data on GDP by industrial origin was largely non-existent. As a result, USAID study relied primarily on observation, anecdotal evidence, press reports and a limited amount of statistical data. According to the USAID (2006, p.13), "…due to the unique and uncertain nature of the current situation, this evaluation is not based upon a completely controllable methodology, but rather upon the best available information, not always reliable."