Journal of Economics and Sustainable Development www.iiste.org
ISSN 2222-1700 (Paper) ISSN 2222-2855 (Online)
Vol.3, No.6, 2012

Analysis of the impact of financial development on Foreign Direct Investment: A Data Mining Approach

Chaitanya Korgaonkar
Department of Information Technology, Veermata Jijabai Technological Institute
(Autonomous institute, affiliated to University of Mumbai) Mumbai – 400019
Maharashtra, India
Tel: 91-9930668966 E-mail: chaitanyakorgaonkar@gmail.com

Abstract
Several empirical studies have tried to examine the links between Foreign Direct Investment, financial development and economic growth. However, little work has been done to examine the direct relationship between FDI and financial development. Thus, through this study we aim to examine if a well-functioning financial system has an impact on the FDI inflows and outflows of a country using the data mining techniques of attribute analysis, association and classification. We have used data related to 78 countries over a period of 1980 to 2009 for our analysis. The analysis suggests that FDI is not directed into countries that are financially weak and is dependent on both the stock market variables and the banking sector variables. The development of the financial system of the recipient country is an important precondition for FDI to have a positive impact on economic growth.
Keywords: Foreign Direct Investment, FDI, Financial Development, Data Mining

1. Introduction

Foreign direct investment (FDI) or foreign investment refers to the net inflows of investment to acquire a lasting management interest (10 percent or more of voting stock) in an enterprise operating in an economy other than that of the investor. This study examines the effects of financial market development on FDI inflows and outflows by studying the link between FDI and the degree of development of the stock market and the banking system. To perform our analysis, we apply the data mining approach.

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information that can be used to increase revenue, cuts costs, or both. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases (Data Mining: What is Data Mining?).

The findings in our work can have important policy implications for various countries in the world. Identifying the relationship can help in improving the quality of local financial systems to make them more attractive for any multinational firms to invest in their markets.

2. Literature Survey

We study the various literatures available in the field of finance, economics and data mining which provide information and a background for our study.

3.1 What determines financial development?

This paper by Yongfu Huang (2005) studies the principal determinants that affect financial development. The author reaches the conclusion that the level of financial development is determined by its institutional quality, government policies, geographic endowments, its income level and cultural characteristics. Of the 39 variables used by in the analysis, 8 variables - initial income, initial population, land area, open trade policy, civil law countries, common law countries, a governance index and a political constraint index were found to be associated with the financial development. More open trade policies are associated with greater financial development and better institutional quality and higher levels of civil liberties and political rights are also associated with higher levels of financial development. The finding that the legal origins influence financial development supports the emphasis on legal determinants of financial development of (La Porta). Findings on institutions, policy and geography as a whole being all important for financial development have significant implications for developing financial markets. The author mentions that efficient supply of external finance can be achieved through good institutional quality. The significant effects of the structural factors which are relatively time-invariant means that any effort of the government to better institution quality implement more open trade and sound macroeconomic policies, and improve geographic infrastructure can stimulate financial development in the long run.

3.2 Foreign Direct Investment, Productivity, and Financial Development

This paper by Niels Hermes and Robert Lensink (2003) argues that a more developed financial system has a positive effect on the process of technological diffusion that FDI brings in. Thus the author states that the development of the financial system is an important precondition for the FDI to have a positive impact on the economic growth. Data from 67 countries was used to perform empirical analysis. Out of these, 37 countries mainly from Latin America and Asia have a developed financial system which helps FDI in contributing to the economic growth of the country. The paper also deduces that the FDI does not have an impact on the economic growth in least developed countries. This means that only when these countries have achieved a significant development of their financial system can FDI have a positive impact on the economic growth.

3.3 The causality relationship between financial development and foreign direct investment

This paper by Zukarnain Zakaria (2007) carries out a systematic study to determine whether a causal relationship exists between FDI and the level of financial development. The author uses data from 37 developing countries for examining this causality in a multivariate framework. The findings from causality tests provide little support for the hypothesis that the inflows of FDI can contribute to the development of the domestic banking sector in developing countries. This study also finds that FDI has no effect on the development of the domestic banking sector. In contrast, the author finds strong support that FDI can affect the development of the domestic stock markets in the developing countries, and vice versa.

3.4 Do Well-Functioning Financial Systems affect the FDI flows to Latin America?

This paper by Omar M. Al Nasser and Xavier Garza Gomez (2009) examines the direct relationship between FDI and the development of the stock market and banking system using the pooled data of 15 Latin American countries from 1978 to 2003. The paper finds that FDI is positively correlated with trading volume (TV), an important variable that reflects the development of the stock market. FDI is significantly and positively correlated with the level of private credit (PC) offered by the banking sector and that the effect of TV and PC is incremental over control variables such as inflation, openness of the economy to foreign trade, technology gap and infrastructure level.

3.5 An Introduction to Variable and Feature Selection

This paper by Isabelle Guyon and Andr´e Elisseeff (2003) provides an understanding of feature, construction, feature ranking, multivariate feature selection, efficient search methods, and feature validity assessment methods. The objective of variable selection is three-fold: improving the prediction performance of the predictors, providing faster and more cost-effective predictors, and providing a better understanding of the underlying process that generated the data. The paper also contributes by providing a better definition of this objective function. The authors recommend using a linear predictor of one's choice and select variables in two alternate ways: (1) with a variable ranking method using a correlation coefficient or mutual information; (2) with a nested subset selection method performing forward or backward selection or with multiplicative updates.

3.6 Feature Subset Selection: A Correlation Based Filter Approach

This paper by Mark A. Hall and Lloyd A. Smith describes a feature subset selector that uses a correlation based heuristic to determine the goodness of feature subsets, and evaluates its effectiveness with 3 machine learning algorithms: a decision tree inducer, a naive Bayes classifier, and an instance based learner. Results show that its evaluation heuristic chooses feature subsets that are useful to common machine learning algorithms by improving their accuracy and making their results easier to understand. A comparison with the Wrapper approach to feature selection shows CFS to be many times faster, making its application to domains with many features more feasible.

3.7 Correlation-based Attribute Selection using Genetic Algorithm

This paper by Rajdev Tiwari and Manu Pratap Singh formulates and validates a method for selecting optimal attribute subset based on correlation using Genetic algorithm (GA), where GA is used as optimal search tool for selecting subset of attributes. Given two attributes, such analysis can measure how strongly one attribute implies the other, based on the available data. For numerical attributes, we can evaluate the correlation between two attributes, X and Y, by computing the correlation coefficient. In general, a feature/attribute is good if it is relevant to the class concept but is not redundant to any of the other relevant features. If we adopt the correlation between two variables as a goodness measure, the above definition becomes that a feature is good if it is highly correlated to the class but not highly correlated to any of the other features.

2. Design Phases

For our analysis, we divide our study into 3 phases. Each data mining technique enables us to get more information about our data and helps us analyze it.

Phase 1: Data Preprocess

We perform data preprocessing in order to integrate different data sets and to clean the missing values. We also apply Discretization to our data when it's necessary for our analysis tasks.

Phase 2: Data Characterize

In this phase, the high level data descriptions are performed. We use Attribute Analysis techniques to get a better understanding of how the data distribution looks like.

Phase 3: Associate Classify/ Predict

We use Association method for discovering interesting relations between various attributes of our interest. We use Classification technique to predict group membership for the different attributes that we have chosen. For that we use popular classification techniques like regression and k-nearest neighbors.

The Datasets used in our study are World Bank Research Dataset, UNCTAD Databases (United Nations conference on Trade and Development Database), International Monetary Fund Dataset (IMF Dataset).

2.1 Data Selected

2.1.1 Countries covered

World Bank Research Dataset includes 224 countries. IMF has total 183 countries with their 40 different financial indicators in Excel file format when United Nations dataset contains around 230 countries. We use the intersection set to select the common data.

2.1.2 Year covered

These three datasets give different time periods for the data. We will basically choose the years which all datasets cover. World Bank Research Dataset addresses the financial indicators from 1960 to 2009. IMF has the period from 1970 to 2005 and United Nations has the data between 1980 and 2009. We thus work on the years from 1980 to 2009 for time period.

2.1.3 Criteria and Indicator Selected

For World Bank Research Dataset, their dataset addresses totally 17 different indicators that measure the size, activity, and efficiency of financial intermediaries and markets. Here we select those indicators we think relevant to international direct investment.

2.2 Data Integration

Since we use three different datasets in this study, we have to deal with data integration. First, these datasets address different topics and indicators. Second, the measurements of them are different as well. We choose the relevant topics and indicators from each dataset and integrate all the indicators into a final combined dataset.

All those countries are identified for which we have very less amount of data i.e. most of the values are missing and they are not considered for our study. Final integrated data is obtained by filling the missing values in the dataset. We replace the empty cell by the mean of that attribute for that particular country.

3. Attribute Analysis

In order to analyze the relationship between the Foreign Direct Investment and the financial development of the country, it is essential to identify the relevant attributes in the dataset.

We use Weka as our data mining tool, it supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection. It provides various attribute evaluation techniques. The attributes are then ranked according to their computed relevance to the data mining task. Attributes that are not relevant or are weakly relevant to the task are then removed. These methods enable us to make a preliminary analysis of the relationship between FDI and financial development and allow us to select only those relevant attributes for a more in-depth data mining.

3.1 Selection of Class Attributes

As we have to analyze the relationship between FDI and the financial development indicators, we select the attributes related to FDI as our classes, and then evaluate the relevance of the financial development indicators to these classes. Thus, the classes selected by us are:

1. Inflows As A Percentage Of GFCF (Gross Fixed Capital Formation)
2. Inward Stock As A Percentage Of GDP (Gross Domestic Product)
3. Outflows As A Percentage Of GFCF
4. Outward Stock As A Percentage Of GDP

3.2 Evaluating and Searching relevant attributes

Subset evaluators take a subset of attributes and return a numeric measure that guides the search. They are configured like any other Weka object. Single attribute evaluators evaluate each of the attributes individually and sort them, discarding those attributes that fall below a chosen cut off point. Search methods traverse the attribute space to find a good subset. For analyzing the attributes, we used two sets of evaluators and search methods.

3.2.1 Feature Selection Method: Correlation based Feature Selection (CFS)

Search Method: Genetic Algorithm

CFS evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Correlation coefficient is used to estimate correlation between subset of attributes and class, as well as inter-correlations between the features. Relevance of a group of features grows with the correlation between features and classes, and decreases with growing inter-correlation (Mark A. Hall). CFS is used to determine the best feature subset and is usually combined with search strategies such as forward selection, backward elimination, bi-directional search, best-first search and genetic search.

Genetic Algorithm is a stochastic general search method, capable of effectively exploring large search spaces, which is usually required in case of attribute selection. Further, unlike many search algorithms, which perform a local, greedy search, GAs performs a global search. A genetic algorithm mainly composed of three operators: reproduction, crossover, and mutation. Reproduction selects good string; crossover combines good strings to try to generate better offspring's; mutation alters a string locally to attempt to create a better string. In each generation, the population is evaluated and tested for termination of the algorithm. If the termination criterion is not satisfied, the population is operated upon by the three GA operators and then re-evaluated. This procedure is continued until the termination criterion is met (D. Goldberg).

The subset of attributes related to the financial structure and development obtained by applying CFS and Genetic algorithm, which have correlation with the FDI inflows to the country are:

Class Attribute: FDI inflow as percentage of GFCF

Country, year, central bank assets to GDP, private credit by deposit money banks, stock market capitalization to GDP, current account balance GDP per capita, GDP purchasing power parity, GDP deflator, inflation annual per capita.

Class Attribute: FDI outflow as percentage of GFCF

GDP, purchasing power parity, life insurance penetration and per capita GDP.

Class Attribute: FDI stock inflow GDP per capita

Deposit money bank assets, GDP deflator, private credit and stock market capitalization.

Class Attribute: FDI stock outflow GDP per capita

GDP purchasing power parity, per capita GDP, life insurance penetration.

3.2.2 Feature Selection Method: Information Gain Analysis

Search Method: Ranker

This selection algorithm evaluates the worth of an attribute by measuring the information gain with respect to the class. It discretizes numeric attributes first using the MDL (Minimum Description Length)-based discretization method. This method can treat missing as a separate value or distribute the counts among other values in proportion to their frequency.

Ranker is not a search method for attribute subsets but a ranking scheme for individual attributes. It sorts attributes by their individual evaluations and must be used in conjunction with one of the single-attribute evaluators. Ranker not only ranks attributes but also performs attribute selection by removing the lower-ranking ones. You can set a cutoff threshold below which attributes are discarded, or specify how many attributes to retain.

Table 1: Ranks of the attributes for their respective classes

Classes / Attributes	Inflows As A Percentage Of GFCF	Inflows As A Percentage Of GFCF	Inflows As A Percentage Of GFCF	Inflows As A Percentage Of GFCF
Country	1	1	1	1
Year	2	12	2	20
Development	20	19	20	18
Deposit money bank versus central bank assets	11	14	14	12
Central bank assets to GDP	14	18	10	17
Deposit money bank assets to GDP	9	10	3	7
private credit by deposit money banks	6	8	5	6
Private credit by others too	10	7	8	5
Liquid liabilities to GDP	13	13	9	11
Life insurance penetration	19	3	19	4
Non-life insurance penetration	16	9	15	8
stock market capitalization to GDP	12	5	6	10
Stock market total value traded to GDP	7	6	17	9
Stock market turnover ratio	15	11	18	13
current account balance GDP per capita	18	20	16	19
GDP purchasing power parity	5	2	11	2
GDP purchasing power parity sow	4	16	7	14
GDP per capita current prices	8	4	13	3
GDP deflator	3	17	4	16
inflation annual per capita	17	15	12	15

3.2.3 Conclusion

The selected attributes from different set of evaluators and searching methods are quite different. Some attributes selected from the first set ranked rather low in the second set. Thus, after filtering the attributes from the 3 combinations of evaluation and selection methods, we find that the most relevant attributes corresponding to the Foreign Direct Investment are:

1. For FDI Inflows:

Country, year, central bank assets to GDP, private credit by deposit money banks, stock market capitalization to GDP, GDP purchasing power parity, GDP purchasing power parity sow, GDP deflator, stock market total value traded to GDP, deposit money bank assets to GDP, inflation annual per capita.

2. For FDI Outflows:

Country, development, private credit by others, life insurance penetration, nonlife insurance penetration, stock market capitalization to GDP,stockmarkettotalvaluetradedtoGDP,GDP_per_capita_curr_prices, private credit by deposit money banks, inflation annual per capita.

3. FDI Inward Stock:

Country, year, central bank assets to GDP, private credit by deposit money banks, liquid liabilities to GDP, stock market capitalization to GDP, GDP purchasing power parity sow, GDP deflator, inflation annual per capita, deposit money bank assets to GDP, private credit by others.

4. FDI Outward Stock:

Country, private credit by others, life insurance penetration, stock market capitalization to GDP, stock market total value traded to GDP, GDP purchasing power parity, deposit money bank assets to GDP, private credit by deposit money banks.

4. Association

In data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases.
We will be using Apriori Algorithm for our study. The name of the algorithm is based on the fact that the algorithm uses prior knowledge of frequent item set properties (Jiawei Han).
We implement the association analysis for the four class attributes separately. First we have to select the attributes those are relevant to the class attributes, this was already done as a part of the Attribute Analysis.
The results of the Apriori Algorithm are as follows.
Class Attribute: FDI inflow as percentage of GFCF
The results show that the inflow of foreign direct investment when described as a percentage of Gross fixed capital formation is closely associated with Purchasing Power Parity and GDP deflector.
Class Attribute: FDI outflow as percentage of GFCF
The results show that the outflow of foreign direct investment when described as a percentage of Gross fixed capital formation is closely associated with the private credits and the GDP per capita current prices.
Class Attribute: FDI stock inflow GDP per capita
The results show that the per capita inflow of stocks is closely associated with GDP deflator and it is associated with other attributes strongly when they are taken in combination.
Class Attribute: FDI stock outflow GDP per capita
The results show that the per capita inflow of stocks is closely associated with private credits and ratio of deposited assets in bank to GDP.
The association rules can be useful for predicting the range in which the FDI can lie given the range of the financial development variables that the FDI variables are associated with.

5. Classification and Prediction

Classification is a data mining technique used to predict group membership for data instances. It is the task of generalizing known structure to apply to new data.
Prediction is similar to Classification. It constructs a model and uses it to predict unknown or missing values. The basic difference however between classification and prediction is that classification predicts categorical class labels whereas prediction models continuous-valued functions.
We have to analyze correlation between the financial development variables and foreign direct investment of various countries. Because variant size of one country's financial structure may have different international investment behavior and strategy, understanding the correlation between the two can be very useful for the economy of a country.
There are various algorithms that can be used to classify the data and further predict the unknown values. In our study, we will be primarily concentrating on Regression Analysis technique (Interpreting Regression).
5.1 Linear Regression
Regression is a data mining technique used to fit an equation to a dataset. The simplest form of regression, linear regression, uses the formula of a straight line (y = mx + b) and determines the appropriate values for m and b to predict the value of y based upon a given value of x. Advanced techniques, such as multiple regression, allow the use of more than one input variable. Multiple regression fits a model to predict a dependent (Y) variable from two or more independent (X) variables: $Y= \beta_0 + \beta_1X_1 + \beta_2X_2 + + \beta_3X_3 + \ldots\ldots + \beta_nX_n$, where, the quantities $\beta_{0\ldots\ldots}$ β_n are unknown coefficients, whose values are determined by least squares.
5.1.1 Selection of attributes for regression
We use the regression model to identify the impact of the various financial development attributes on the FDI data.
Originally we have 38 attributes in our dataset. After using all of those attributes to do the mining task, we found that due to enormous variation of some values of the attributes, the resulting prediction is not very satisfied, even though we have done attributes normalization to eliminate the problem stemmed from scalability. So we decide to filter out those attributes whose values are absolute and varying dramatically from country to country. We keep only those attributes which are in percentage or ratio.
Secondly, we remove country name, year, and development attributes. Because our goal is to build a model which can make a prediction on one nation's inward or outward investment based on other attributes' values excluding country, year and development situation. Therefore, we select the attributes based on the results of attribute analysis discarding the three attributes of country, year and development.
1. FDI stock outflow as a percentage of GDP
Attributes selected:

Private credit by others, life insurance penetration, stock market capitalization to GDP, stock market total value traded to GDP, GDP purchasing power parity, deposit money bank assets to GDP, private credit by deposit money banks

Analysis:

Here we find that only 2 variables are significant. The stock market total traded value as well as the GDP purchasing power parity has a positive and significant impact on the FDI inflows.

2. FDI outflows as a percentage of GFCF

Attributes selected:

Private credit by others, life insurance penetration, nonlife insurance penetration, stock market capitalization to GDP, Stock market total value traded to GDP, GDP per capita current prices, private credit by deposit money banks, inflation annual per capita.

Analysis:

Here 5 variables are significant. The life insurance penetration variable negatively and significantly impacts the FDI outflows while the non-life insurance penetration is positively correlated to the outflows. The stock market capitalization and the total traded value also have a significant impact on the outflows. Thus, the stock market development in terms of total market capitalization as well as the total traded value affects the FDI outflows.

3. FDI stock inflow as a percentage of GDP

Attributes selected:

Central bank assets to GDP, private credit by deposit money banks, liquid liabilities to GDP,
Stock market capitalization to GDP, GDP purchasing power parity sow, GDP deflator, inflation annual per capita, deposit money bank assets to GDP, private credit by others.

Analysis:

Here we find that 6 variables are significant. The central bank assets and the GDP purchasing power parity variables seem to have a positive impact on the FDI inward stock. The inflation annual percentage variable also has a significant impact on the FDI. The deposit money bank assets variable has a negative impact on the FDI inward stock.

4. FDI inflows as a percentage of GFCF

Attributes selected:

Central bank assets to GDP, private credit by deposit money banks, stock market capitalization to GDP, GDP purchasing power parity, GDP purchasing power parity sow, GDP deflator, stock market total value traded to GDP, deposit money bank assets to GDP, inflation annual per capita

Analysis:

Here 7 out of the 8 variables are significant. This shows that these variables have a significant impact on the FDI outward stock.

After running regression analysis on the data we found out that there some of the financial indicators have an impact on FDI. To further check out if the FDI can be predicted for a country based on the financial development indicators, we applied few more algorithms. The best results were shown by K-Nearest Neighbor algorithm which is a distance based classification algorithm.

5.2 K-Nearest Neighbors (KNN)

K-Nearest Neighbors is a common classification scheme based on the use of distance measures. The KNN technique assumes that the entire training set includes not only the data in the set but also the desired classification for each item. In effect, the training data becomes the model. When a classification is to be made for a new item, its distance to each item in the training set must be determined. Only the K closest entries in the set are considered further. The new item is then placed in the class that contains the most items from this set of K closest items (Margaret H. Dunham).

5.2.1 Selection of Attributes:

As reasoned out for regression, it is inappropriate to apply the algorithm for all the attributes. Thus, the relevant attributes are selected based on the results of attribute analysis and the algorithm is applied only to these selected attributes.

Analysis:

1. FDI stock outflow as a percentage of GDP

On applying this algorithm, we find that the coefficient of correlation is high and the relative absolute error is low. This suggests that the K-Nearest algorithm can be used to predict the FDI outward stock.

2. FDI outflows as a percentage of GFCF

Here, we find that the coefficient of correlation is not high enough which means that there is significant difference in the predicted value and actual value of FDI outflows. Thus, the K-Nearest algorithm is not fit for predicting the FDI outflows.

3. FDI stock inflow as a percentage of GDP

The coefficient of correlation is very high for the FDI inward stock and the relative absolute error is sufficiently low. This suggests that the K-Nearest algorithm can be used to predict the FDI inward stock with greater accuracy.

4. FDI inflows as a percentage of GFCF

Here, we find that the correlation coefficient is sufficiently high. Thus, the K-Nearest algorithm can be used to predict the FDI inflows.

5.3 Conclusion

Linear regression techniques allowed us to analyze the impact of various financial development indicators on the FDI. Also K-Nearest algorithm can be used for predicting the FDI inflows, inward and outward stock with .least error.

6. Conclusion

As discussed in the previous chapters, we have applied the data mining approach for analyzing the impact of the financial development variables on the Foreign Direct Investment to and out various countries. Feature Selection methods provided preliminary analysis about the financial development variables that can be useful for predicting the amount of Foreign Direct Investment. Association rules on applying the Apriori algorithm provide us with various rules which allow us to predict the range in which the FDI values can lie given the range of the financial development variables to which it is associated. We measure stock market development by stock market capitalization as a percentage of GDP, total value traded to GDP, and the turnover ratio. Through regression analysis, we find that stock market capitalization is positively and significantly correlated to the FDI inward stock and inflows to the countries. The stock market development in terms of stock market capitalization and total value traded is also a strong predictor of the FDI outflows. The banking sector development variables in terms of central bank deposits and deposit money bank assets variables have a significant impact on the FDI inward stock and inflows to countries. The central bank assets variable is positively correlated to both while deposit money bank assets variable is negatively correlated to both. We find that the K-Nearest algorithm fits the data best to predict the FDI inflows, inward and outward stock with less error as compared to other models.

Thus overall the analysis suggests that FDI is not directed into countries that are financially weak and is dependent on both the stock market variables and the banking sector variables. The development of the financial system of the recipient country is an important precondition for FDI to have a positive impact on economic growth.

References

D. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison Wesley,(1989).

Data Mining: What is Data Mining?
http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm

International Monetary Fund
http://www.imf.org/external/data.htm

Interpreting Regression
http://dss.princeton.edu/online_help/analysis/interpreting_regression.htm

Isabelle Guyon, Andr´e Elisseeff (2003), "An Introduction to Variable and Feature Selection", Journal of Machine Learning Research 3 (2003), 1157-1182

Jiawei Han and Micheline Kamber, "Data Mining- Concepts and Techniques", 2nd edition

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, R. W. Vishny (1998), "Law and Finance", Journal of Political Economy, 106: 1113-1155.

Margaret H. Dunham, "DATA MINING: Introductory and Advanced Topics"

Mark A. Hall, Correlation-based Feature Selection for Machine Learning, Department of Computer Science, University of Waikato.
http://www.cs.waikato.ac.nz/~mhall/thesis.pdf.

Mark A. Hall, Lloyd A. Smith, "Feature Subset Selection: A Correlation Based Filter Approach", Department of Computer Science, University of Waikato, Hamilton, New Zealand.

Niels Hermes and Robert Lensink (2003), "Foreign Direct Investment, Financial Development and Economic Growth", Journal of Development Studies, Vol.38

Omar M. Al Nasser, Xavier Garza Gomez (2009), "Do Well-Functioning Financial Systems Affect the FDI Flows to Latin America?", International Research Journal of Finance and Economics, Issue 29

Rajdev Tiwari, Manu Pratap Singh, "Correlation-based Attribute Selection using Genetic Algorithm", International Journal of Computer Applications (0975 – 8887), Volume 4– No.8, August 2010

United Nations conference on Trade and Development (UNCTAD)
http://unctadstat.unctad.org/ReportFolders/reportFolders.aspx?sCS_referer=&sCS_ChosenLang=en

World Bank
http://www.worldbank.org/

Yongfu Huang (2005), "What determines financial development", University of Bristol

Zukarnain Zakaria (2007), "The causality relationship between financial development and foreign direct investment", Jurnal Kemanusiaan bil.10, Disember 2007