INTERNATIONAL JOURNAL OF RESEARCH & METHODOLOGY IN SOCIAL SCIENCE

VOL. 1, NO. 2 | APR. – JUN., 2015 | QUARTERLY JOURNAL | ONLINE

Scientia, Fortitudo & Victoria

QUATERLY JOURNAL
Online Publication

Vol. 1 No. 2

Peer
Reviewed
Scientific Journal
in Social Science

OPEN ACCESS

OFFICE
OF
THE EDITOR-IN-CHIEF
420 Bourichane Rd.
Unit 1 Ban Naxay
Xaysetha, Vientiane LAOS
Tel. +856 21 263585

IJRMSS

ONLINE & OPEN ACCESS
www.socialsciencepublication.com

OFFICE OF THE EDITOR-IN-CHIEF
International Journal of Research & Methodology in Social Science (IJRMSS)
420 Bourichane Rd., Unit 1 Ban Naxay
Xaysetha, Vientiane
LAOS
Tel. +856 21 263585
Editor@socialsciencepublication.com
EditorIJRMSS@gmail.com
www.socialsciencepublication.com

THE BOARD OF EDITORS

INTERNATIONAL JOURNAL OF RESEARCH & METHODOLOGY IN SOCIAL SCIENCE (IJRMSS) is a peer reviewed scientific journal in social science and its interdisciplinary fields. The journal is published quarterly. The Boards of Editors oversee the acquisition, review, and publication of quarterly research papers in social science. Editorial inquiries may be sent by email directly: Editor@socialsciencepublication.com or EditorIJRMSS@gmail.com.

NOTA BENE

Issues No. 1, No. 2, No. 3 & No.4 of volume 1 for the year 2015 were published and saved in different file format prior to the assignment of the ISSN. These older electronic file versions were damaged during conversion. Damaged electronic files had recently been recovered and saved in PDF format. All issues in volume 1 had been reproduced in PDF format for archiving and retrieval purpose. The current location for the online achieved is found at: http://www.socialsciencepublication.com/ARCHIVE.htm

TABLE OF CONTENT

The Role of family Business in the Economy

Louangrath, P.I. ★

About the author

★Louangrath, P.I. is an Assistant Professor in Business Administration at Bangkok University, Bangkok, Thailand. He could be reached by email at: Lecturepedia@gmail.com

ABSTRACT

The scope of paper is the role of family business in the economy. This topic is important because family business is the platform from which most SMEs sprung and a significant part of the GDP in many countries, particularly those of the ASEAN, depends on SME and family businesses. We assert that family business is a form of direct domestic investment for the economy. Success in family business means economic growth for the nation. As such, the success or failure of the family business serves as an indicator for the health of the national economy. Family business springs from successful entrepreneur; since an entrepreneurial is the source of innovation, family business also serves as an incubator for innovation which, in turn, contributes to economic growth.

Keyword: entrepreneur, entrepreneurship, family business, direct investment, innovation

CITATION:

Louangrath, P.I. (2015). "The role of family business in the economy." *Inter. J. Res. Methodol. Soc. Sci.*, Vol. 1, No. 2: pp. 1-17. (Apr. – Jun. 2015).

1.0 INTRODUCTION

The objective of this paper is to provide an over all view of the role of the family business in the national economy. Family business plays a significant role in the life of the economy. It is a mistaken idea to think that large corporations are the only maker and shakers in the national economy. Large corporations have their humble beginning. They are the out growth of SMEs that had gone through the natural growth process where formal structure becomes institutionalized. SMEs too are the extension of the success of family businesses. In part 1, we have learned that the family business is the development of a successful enterprise of an entrepreneur. This paper paints the picture of the family business as an active participant in the economy. The significant contribution by the family business to the national economy is an indicator of the economic health of a nation. No matter how advanced is the economy family business always plays a significant role in economic growth and job creation for the community. For the state, family business remains a stable source of tax revenue and an important base for political support.

This paper also attempts to answer these questions: *Is there a theory governing or explaining family business? What is a family business model?* The theory of firm explains the omnipresence of the need for profit maximization; however, in family, business where the structure is less formal and the fiduciary is not as stringent as in a formal corporate structure where the stakeholders include people and institution who do not have direct contact with the enterprise, family business deals with a different demographic of stakeholders: family members, customers, suppliers, distribution channels, and, inextricably---the state. Under these facts, *what is the functional model that helps explain the operations and success of a family business?*

2.0 FAMILY BUSINESS, INNOVATION, AND ECONOMIC GROWTH

Family business refers to a business or a commercial enterprise that is closely held or owned within the family. There is no need to define the term family; it is suffice to accept a broad concept that for purposes of family business, the term 'family' refers to the filial relationship that defines the ownership of the business. Where applicable, the terms 'business' and 'enterprise' may be used interchangeably.

Innovation, as defined in paper 1, refers to the increase in economic output with the same level of input (performance innovation) or maintaining the same output with the decrease in input (input innovation) through the improvement of methods or procedure of the production process. Family business is the master in resource management. Family business, like all organization but more so in family business, is faced with limited resources. Family business uses the available resources to create value. The position taken by this book is *value creation*, not *value additivity*. Value creation refers to the creation of value from an existing resource which is not used and, thus, creates no value in a case of an idled asset. In a case where the asset is not idle, the resources are diverted or converted to other activities or the same activity but under a new and improve procedure to create additional output. This process is known as value creation. Value additivity, on the other, is to add value to an existing product. The addition of value is achieve through adding addition features or function into an existing product, as the result cost increases. With the increased cost, in order to maintain the same level of profit the producer has to increase price. The increasing in rice may draw ire from the market if the introduction of the 'improved' product is not carefully orchestrated.

Economic Growth refers to the growth of the general wealth in the economy. This growth is generally measured in a form of the Gross Domestic Product (GDP). Family business is an indispensable part of the economy. As a significant player in the creation of value, and hence, the national wealth of a country, family business serves as an important indicator for the health of the economy. Both foreign and domestic investors may look to the performance by family business as an indicator for the over all health of the economy and make investment decision accordingly. Since family business plays a significant role in contributing to the economic growth of a country, this paper also explores some state policy issues that may be relevant to family business in the section: *Domestic Direct Investment (DDI).*

3.0 THE FAMILY BUSINESS

Family business is a business or enterprise that is owned by one or more members of the family. It is more organized than an enterprise run by an entrepreneur because it is one stage further in development compared to an entrepreneurship. An entrepreneur's enterprise is a risk taking activity in a sense that the entrepreneur answer to no one. Being his own boss, an entrepreneur is a one man show. As such an entrepreneurship is a risk taking enterprise. As a risk taking enterprise, an entrepreneur wage in at all front and make himself appear to have a specialty affinity for risk. In deed an entrepreneur takes risk and entrepreneurship is a risk taking enterprise. To that end, Cantillon was correct in asserting that an entrepreneur, hence an entrepreneurial enterprise, pays a known sum for an unknown gain. A family business, on the other hand, is more formalized and more structured.

A family business is more formalized than an enterprise run by an entrepreneur. The formal structure of a family business lies in the fact that it has moved from a one-man enterprise to an operation that involved more people. These people include family members and other stakeholders, such as employees, suppliers, distributors, and

Other writers claim that as part of the formality, the family business institutes something akin to a 'council.' Although there is a semblance of formality in a sense that the operation involves more than one man, the entrepreneur, a claim of 'council' or a family 'board' is a fantasia. The head of the family business remains the original entrepreneur who started the business. The decision making process remains unidirectional. The information flow and decision making remained unilateral. There might be discussion or consultation among family members; however, ultimately the decision making remains dictatorial. The idea of a family council is a mental construct by writers who try to make sense of the semblance of formality in what otherwise would have been unstructured.

Structure is an additional feature that appears in a family business. When an entrepreneur operates as a sole proprietor, the enterprise does not have any organization and, therefore, is unstructured. Once the entrepreneur transforms the business into a family business, an organizational structure begins to appear. The structure may appear in a form of shared ownership, shared responsibility, shared risk, and share benefits. The ideas of 'shared' interests and responsibilities increase the layers of stakeholders in the enterprise. The decisions were made instantaneously prior to the introduction of shared interests and shared responsibility would now take longer time and involve more people. The conduct of the family business becomes less *ad hoc* and more formal within a confine of well define structure.

The interest of the family and the interest of the business may not coincide. The family may need money for personal expenses, but the business may require the capital to remain in the business account (Loewen, 2008). The formalization and structure in the family business does not necessarily result in complete coherence. In an entrepreneurial enterprise, the business and personal interests are inseparable because the entrepreneur is the sole proprietor. However, in a family business, although the ownership remains within the family, the ownership of the business may vest in more than one hand. As such, the use of any money in the family business may involve more than one decision makers. Sometimes this decentralization of power may cause internal conflict among group members.

The internal conflict may threaten the life of the family business unless it is resolved. Since family business involves mostly members of the same family, conflicts are considered internal and could be resolved internally. Under such circumstance, it is expected that conflict resolution in a family business may be swift because it is not contentious. This is not to say that the idea of 'family council' exists as a means to resolve internal conflict in a family business because ultimately if the conflict festers, the head of the family business will exert authority and assumes the position of a final arbiter. Although the term 'arbiter' is used, there is no arbitration. The decision is more dictatorial then consultative. Consultation implies the dispensation of knowledge by one who knows more to another who knows less about a certain subject matter. However, this type of knowledge seeking does not exist in a family business. The erroneous claim that a family business has a 'family council' comes from that fact that in case where there is an internal conflict, the group attempts to resolve the issue. However, the resolution does not come from consultation. The decision making authority is still retained by the original entrepreneur, i.e. sole proprietor. Internal conflict occurs when the sole proprietor is indecisive and allows the decision making process to be pluralistic. Where there is chaos, the sole proprietor restores order by reestablishing himself as the ultimate decision maker. This decisiveness breaks the indecision from conflicts within the group. The concept of 'democratic decision making' in a family business is a mental construct of academics who failed to fully credit the sole proprietor or entrepreneur of the family business as the final arbiter of all decisions.

3.1 Family Business as a Source of Innovation

A family business may be a source of innovation. Innovation has been previously defined as the use of existing resources to increase productivity. Five elements must present in order to be successful: (i) defined goal, (ii) align goals to actions, (iii) team participation, (iv) communication, and (v) access to information (O'Sullivan, 2002).

Defined goals exists with both entrepreneurial enterprise and family business, but more so in family business. An entrepreneur lives much of his life in the Cantillon world: known fixed cost and unknown income. However, the world of the family business is more finite. The definiteness of the family business comes from the fact that a family business owner involves more people in the enterprise. For that reason, the action of the family business must be guided by set goals or objectives. With definiteness in goals or objectives, the family business builds for itself a confine within which the enterprise operates. The *advantage of defined goals* allows the family business to stay focus of a certain path in order to achieve an increase in economic productivity for any given level of input. A business focus allows the family business to confine its conduct within a certain core areas. Since it involves many people in the enterprise, the decision to stay within a certain confine of business scope is less arbitrary and more thoughtfully planned. Even if the family business owner may at times be dictatorial in the decision making process, the overall condition under which the decisions are made is less arbitrary in the family business setting than it is in the entrepreneurial enterprise run by a single entrepreneur who answer to no one. Moreover, the defined goals of the family business allows the enterprise to contain risk by not over being expansive in business scope. An entrepreneur takes any thing that comes his way---anything that seems profitable. For an entrepreneur, any risk is worth taking so long as the risk promises some degree of returns that exceeds his initial investment. However, a family business cannot afford to have such a liberal approach to resource utilization.

The *disadvantage of defined goals* is that the goal may be over ambitious. Since the family business faces the first barrier in an entrepreneurial enterprise, that is a structural constraint, it cannot afford to be every where at any time. It needs to confine itself to a certain field of business, within a territorial confine, industry, or product line. Unlike an entrepreneur who may search for opportunities and accept every opportunity that comes his way, a family business lacks that liberty due to the structural constraint of the business. Being a family business means that the enterprise has some semblance of an organization. An organization by definition is the institution of structure. Structure means constraint. Constraint means the lack of complete liberty. With defined goals, the family business may not pursue all business opportunities; it could and would pursue only those opportunities that falls within the family business' goals. These goals are determined by the core competence of the family business. The defined goals limit the family business to roam freely in the market. This lack of freedom to roam may be disadvantageous to the family business.

Align goals to actions requires the family business to channel its resources to the undertaking that would lead to goals achievement. The need to achieve in the family business and in an entrepreneur is not the same. An entrepreneur needs to succeed. Success for an entrepreneur is an increase in wealth no matter from what source. However, the family business adopts a set of goals. Resource utilization is dictated by the objectives of these goals. The goals are defined and determined by the family business' area of concentration or core business.

By aligning resources to goals, the conduct and actions of the family business is more concerted. A concerted effort guided by goals and objectives allows the family business to be more organize. Resource utilization by the family business is more effective and efficient. *Effective* means that the resources used to produce a defined target indeed had achieved the production regardless of how much it cost. *Efficiency* is defined by achieving a target at the lowest cost possible. Under these standards, the family business is efficient in resource utilization. The innovative nature of an entrepreneur makes him resourceful through the re-use of existing, but idled, resources.

In goal-action alignment, the family business breaks a new frontier in capacity limitation. An entrepreneur is limited to one man's ability. A family business is more organized, planned, and approach the business in a more structured fashioned. In so doing, the family business can always predict the outcome of the investment. For this reason, the family business frees itself from the confined built by Cantillon: known cost and uncertain income that is characteristic of an entrepreneur. Since the family business has a defined goal and aligns its action in a direction towards achieving that goal, the family business operates in an environment that has a higher degree of certainty.

Team participation exists both in entrepreneurship and in family business; however, team participation in these two businesses setting is not the same. For an entrepreneur, lacking organization he does not have a team; he acts by and form himself. In a family business, the enterprise conducts itself by concerted effort of group members. The head of the family business must be a team player in order to move things along. However, when conflict occurs or there is indecision, the head of the family business must show his absolute authority to break the inaction of the group. There may be grumbling disagreement among some group members, but the organization moves forward and members of the group would learn to accept the reality of a family business: *despite a semblance of structure, the final decision making still rests on the hands of the business owner*.

Communication is defined as "… any act by which one person gives to or receives information from another person about that person's needs, desires, perceptions, knowledge, or effective states. Communication may be intentional or unintentional, may involve conventional or unconventional signals, may take linguistic or nonlinguistic forms, and may occur through spoken or other modes."(*See* National Joint Committee for the Communication Needs of Persons With Severe Disabilities, 1992). In family business, intra-organization communication is more evident that in the case of an entrepreneur who acts alone. By communicating with members of the organization, family business can maintain effective channel for information flow from outside into the organization, from the organization to the outside, and within the organization among members. This open channel allows the family business to gather and evaluate information prior to making decisions. Decision making is not left to chance or arbitrary response. Open channel communication allows the family business to make informed decision. Informed decision is defined as the decision making resulted from taking all available and relevant facts into consideration. Effective communication allows the family business to reduce risk by reducing uncertainty.

Access to information makes innovation possible because information allows the business owner to know the current market condition and anticipate future condition. A family business is a source of innovation when it has full access to information. There are two types of information: useful information and useless information. Useful information must have the following features: *relevant, timely,* and *accurate*. An entrepreneur is at disadvantage when it comes to access to information because by operating in an unorganized environment and without structure, he is left to his core competence. The information that is available to him is circumscribed by his vision, which sometimes may be myopic. A family business, on the other hand, has better access to information because a family business acts as a group. Members of the group may contribute to the sources and access to information. In so doing, the decision made in a family business is more informed; therefore, a family business has less exposure to risk from misinformation or making decision without adequate information than its entrepreneur counterpart.

4.0 INNOVATION AS A SOURCE FOR ECONOMIC GROWTH

Economic growth may be seen at two levels. At the micro level, growth may be evidenced by the profitable operation of the family business. At the macro level, growth may be seen as the aggregate increase of the economic production of the country. The family business being part of the economy plays a role in contributing that growth. Several questions are presented: *What is the role of innovation in family business? How does innovation contribute to growth of the economy?*

Innovation and family business are connected by the extension of success and development of an entrepreneur whose enterprise has outgrown the one-man operation. As an individual, and by definition of 'entrepreneurship,' the entrepreneur is resourceful and innovative. He is resourceful because he uses idled assets into a productive source of income. He is innovative because an entrepreneur can use the existing resources, idled or not, and produce something new, better and more. This trait of the entrepreneur is carried into the family business. Thus, one source of innovation in the family business comes from the entrepreneur who originated the family business. This is called the *spill-over effect*.

In addition to the spill-over effect, the family business may instill into itself the innovative spirit from within the organization. Family business by its nature is small in size, although some family business may be large and dominate an industry, in order to fully understanding the inner working of a family business it is imperative that we work with a simpler model: a small family business unit. In a small family business unit, the unit is left to fend for itself. Corporations have better access to the market and the institutional support in the market, such as banking and finance. However, a family business does not have that luxury or opportunity due to its small size and informal structure which---according to the formalized market structure---lacks creditability and market standing. Therefore, the family business is left to fend for itself.

By fending for itself, a family business is always take and offensive and defensive stands. Offensively, the family business must break new ground in the market in order to stay alive. The primary goal is to stay alive. The ideas of go ahead and stay ahead for family business is still a distant ideal. In order to stay alive in the market, in face of growing competition and the increasing squeeze from larger and more formidable corporation, a family business is compelled to be even more resourceful than an entrepreneur. By enacting self-defense mechanism, a family business searches for new means of doing the same thing and to do new things with the limited resources that it possesses. In so doing, the family business becomes innovative.

Innovation is the application of the current resources to increase productivity. It is the doing the same thing with the same resources, but producing a higher level of output. An innovative family business is always under pressure to produce more. In order to produce more, with limited resources under its control, the family business must be resourceful. If innovation means efficient production and efficient production means increase the production level with the same level of input, then it is reasonable to posit that the family business contributes o the increase in economic growth. Family business is a significant component of the economy; therefore, this argument is logical and is not a leap pf faith. In some countries, family-owned SMEs may account for as much as 40% of the total GDP.

The second question asks: *How does innovation contributes to economic growth?* Innovation means an increase production with the same input level. Another definition of innovation is efficiency. There are two types of innovation: input innovation and output innovation. Input innovation refers to cost reduction in order to maintain the same level of production. Output innovation refers to the increase of production level with the same level of input. Under these definitions, it is clear that where there is innovation there involves the increase in production, i.e. economic growth.

Economic growth had been explained by many writers. Classical economist explained economic growth as the result of capital accumulation; however, in the 21st century the new theory is *innovation economics*. According to innovation economics, economic growth results from innovation and spur in technological advances (Antonelli, 2003). Under this new theory, the number of patent applications is used as evidence to prove that innovation is the driver for economic growth (Fornmahl *et al*., 2011). Economic growth does not come from capital accumulation; growth comes from the close link between business and technology (Peilei, 2011). The steady growth of the US economy was a result of relentless pursuit of technology (Steil *et al*., 2002). Economic growth comes from innovation (Ahlstrom, 2010), not capital accumulation as classical economist had preached. This new theory is particularly relevant to family business because at the heart of family business is the entrepreneurial spirit that spilled over from the metamorphosis of an entrepreneur to a family owned business. As such, the innovative spirit remains active in the family business. Success in family business, in face of keener competition from more formalized and structured corporate producers, is due to its innovative capacity. To that end, because family business constitutes a significant part of the national economy, family business plays a role in economic growth. Since growth comes from innovation---at least under innovation economics school of thought; therefore, the contribution to the over all economy by family business is a further evidence to support the new school of though.

5.0 THE ROLE OF THE FAMILY BUSINESS IN THE ECONOMY

Among listed international firms, 45% are family owned (La Porta and Schleifer, 1999). The family business always has a role to play in the economy. It is central to the inner function and working of the national economy. By definition, family business starts out small and many also remain small, but active in the economy. Since family business requires less capital to operate, the barrier to entry into this type of operational and ownership structure is less restrictive. Therefore, the potential for sector growth is big.

If the life cycle of a business is examined, it is possible to see that family business has a significant role to play in the economy. In the most simples form, an individual entrepreneur engages in an enterprise. Assume that the enterprise becomes successful and grows this growth would naturally lead to the inception of a family business. Not all successful entrepreneurs involve their families and not all of them transform themselves into a family business. There are some successful entrepreneurs who remain unorganized and uninvolved. The alternative route is the second stage where the successful entrepreneur involves his family. In so doing, family members become involved n the enterprise. The enterprise, unlike in the hands of the entrepreneur where structure and organization did not exists, the family business is organized and structured. The size of the operations may also expand to a larger size under the family ownership. The scope of the business may more be focused and risk exposure is reduced. In time, if the family business continues to succeed, it may enter the next stage: SMEs and will ultimately become a corporation. At every stage of the development, the business generates economic activities. These activities contributes to the over all economy.

5.1 The Role of the Economy in Family Business

The previous section speaks to the role of the family business in the economy. The economy also has a role to play in the family business. The operations, i.e. success and failure of family business depend on its environment. That environment is the over all economy. The role of the economy in family business is less explored. This aspect of the study is often overlooked because the economy is a macro picture. As a macro factor, we tend not to consider its role in small business. However, small business operates inside this larger environment. If the health of the economy is ailing, the operations of the family business may also be affected.

Too look at only the role of the family business in the economy is to paint an incomplete picture of the state of affairs. The economy also has a role to play in the family business because the family business operates in the greater operating environment, i.e. the market and the market is the economy. As stated in the previous section, the growth of the economy also depends on the growth of family businesses. If the economy slows down, the family business will be adversely affected.

The owner must take this point into consideration in the management of the family business. This point becomes more relevant in the context of growth and development. Family business is not the ultimate developmental climax of a business. A successful family business will grow into SME and from SME the business will reach a corporate structure. As the economy grows, the family business would also grow with it; with that growth, it will enter a new stage of its life cycle. As the economy slows down, this too will also affect the family business.

The question then arises: *in time where the economy recedes, can family business regress to an unstructured enterprise, i.e. unorganized entrepreneurship?* In the worst case scenario, that would be the natural trajectory of the business. If the adverse effect from the economic recession forces the family business to scale down, the natural path of the shrinkage is to revert to being an unorganized enterprise run by an entrepreneur. This directional movement is worth exploring because not too many type of organization can be such flexible. When a corporation fails, it shuts down and goes out of business. However, when a family faces hard time, it can scale down and ultimately survive by its bare bone form: entrepreneurship. Entrepreneurship is a fall back position for the family business. However, for a corporation failure leads to bankruptcy and ultimately dissolution. To that end, the family business is more flexible in structural adjustment.

The role of the family business is more enduring that the corporation. A failed corporation has no where to run or hide. A corporation moves in one direction. A corporation does not have too many choices; there is no fall back position if it faces an enduring recession. The family business, on the other hand, retains the flexibility of having entrepreneurial enterprise as a cushion against external economic shock. Under this perspective, the role of the economy in the family business is vital, but not fatal. In times of success and growth, the family business rides along on the rising wave of that growth. In time of recession, the family business may scale down to its bare bone origin of entrepreneurship. As the economy recovers, the entrepreneur will pick up the operation and continue his business. The cycle continues. The economy exerts its forces upon the family business. The family business is in a response mode when the economy is bad. In time of growth, the family business may be on the proactive mode in putting itself forward in the market and expand its operation to propel itself to the next stage of development: SME and ultimately a corporation.

5.2 Family Business and Direct Investment

The term direct investment is usually referenced to "foreign direct investment" or FDI. However, in this section, direct investment refers to the fact that in family business, since the structure of the ownership is retained within the family, the investment generally comes from the family fund. The capital structure of the family business is not as complicated as common found in the corporate setting. Family business and direct investment go hand-in-hand. Generally, family business grows out of the success of an individual entrepreneur who then involves other family members in the enterprise. For that reason, the capital used in the family business generally consists of funds that are pooled from family members. In such a closely held situation, there are both advantages and disadvantages. This section explores the source of capital as a source of direct investment in the family business and how pooled fund among family members contribute to potential conflict within the group.

Pooled fund is a form of joint contribution among family members. The money is collected from family members who contribute to the family business. The contribution to the pooled fund may be in a form of loan or direct investment. As a loan, the family business would have to repay the loan. As a pooled fund for direct investment, generally the owner of the money becomes involved in the family business, such as a employee of that business. In both situations the pooled fund is raised outside of the conventional banking channel. For this reason, family business is often looked upon as an *informal sector* of the economy.

Family business is considered part of the informal economy because its capital is raised outside of the conventional banking system. In addition, the ownership of the business is closely held and the size of the operation is generally small in comparison to other forms of businesses, i.e. those who had succeeded to SMEs or corporation. Moreover, the structure of the family business is also informally organized. The ownership, control, and operation of the business are kept inside the family. Lacking general public participation in the ownership structure and small size, family business remains part of the informal sector.

5.3 The Role of Family Business in the Economy *Revisited*

Despite the fact that family business is part of the informal sector, family business plays a significant role in the economy. For that reason, the performance of family business may be used as the indicator for the economic sentiment in the industry, sector or the economy as a whole. In most developing and undeveloped nations, family business comprised a larger share of the economy. In the advanced and more developed economies, family business also plays a significant role in the economy.

In addition to making contribution to the economic production of the national economy, family business also indirectly provides important information to outsiders. Foreign investors looking for investment opportunities in a country may assess the economic environment of the country by looking at how family business and the rest of the informal sector's performance as part of the decision making process. If family business is succeeding, larger corporation may see unfulfilled needs in the market and might considered such a market as a target for penetration. Moreover, failure of family business also provides useful information to foreign investors. The failure of family business may suggest that the business environment is competitive and not conducive to small scale operation. There are many possible conclusions that could be drawn from the condition and status of family business.

Family business is a useful indicator for foreign investment. Foreign investment in a country generally does not come directly to family business. Foreign investors do not seek partnership among family businesses. However, they use the performance of family business as part of the decision of whether to penetrate the market or not. Family business is a risk indicator for foreign investors. This beneficial role of family business has never been recognized. It has been overlooked because the general economy still treats family business and part of the informal sector.

The perception of family business as part of the informal economy also limits the growth and pace of change among family business. Since family businesses are generally small and possess smaller amount of assets, they not as attracting to the leaning market. This perception further contributes to the limitation of expansion as the result of lacking access to the capital market. Consequently, the capital structure of a family business is inward looking and limited to the availability of personal fund among members of a small circle.

The concept of pooled fund among family members forces family business to be more conservative in the use of the fund. This limited fund further forces the family business to be financially disciplined in selecting the field of its investment, direction of business expansion, and in general its resource utilization. In so doing, the family business is forced to adopt more stringent rule in resource allocation and risk exposure. His is part of the process that transforms the family business from an informal and less structured entity to the SMEs with more formality and structural constriction. *Under such circumstances, what is the role of the entrepreneurial spirit among the family business owner?*

As the family business succeeds and develops, and the structure of the business becomes even more formalized, the family business moves to the next stage of development: SMEs. The entrepreneurial spirit is put under the constraint of the organizational structure. Formalization means the lost of liberty. The liberty to think and explores is further constrained by the organizational structure and rules. The entrepreneurial is not dead but remains dormant until it is awaken. The entrepreneurial spirit will be awakened when the family business is threatened with failure or competition. The topic of SMEs is reserved for Part 4 of this 4 parts series.

5.4 Contributing Factors for Direct Investment

The positive performance of family business in a larger picture has a positive effect on foreign investment. This fact is less talked about. No foreign companies want to go into a country and invest where there are many failures among companies in the informal sector, i.e. family business. Family business is a contributing factor to foreign direct investment (FDI).

Why family business has never been recognized for this role it plays in the national economy? In the US, the government general pays attention to the health of the economy. In so doing, the government also uses the grass root economy at the local and state level to assess the health of the national economy. Large corporations failed in the US. However, the failure of these large corporations generally disappear from state policy the soon the bankruptcy proceedings are over. However, the failure of small businesses, such as SMEs or family owned business reverberate more among the American policy makers because these are the people who holds the political mandate for the next election. For that reason all presidential candidates solicit the support from small businesses.

5.5 Investing in Family Business

The conventional definition of investment is the putting of money into a project or enterprise with the expectation of return of the initial capital invested plus profit. Investment in family business occurs in two contexts: (i) the money put into the family business by the stakeholders is considered an investment, i.e. the pooled fund is a direct investment; and (ii) the infrastructural provision by the government through state policy is a form of indirect investment.

Direct investment is the putting of cash or kind which could be valued in money into the project or enterprise of the family business. The pooled fund is the most common form of direct investment. The family business owner may borrow money from outside of the family circle and uses that money for the business operation. This loan is treated as a liability in accounting; however, in economic term, it is an investment because the use of that money is the utilization of resource for the purpose of making more money. However, direct investment in family business is not attractive to investors because family business is in the informal sector. As such the family business does not attract many outside investors. Moreover, formal lending institutions traditionally see family business as a high risk client. It is not until recently that banks in developing countries see family business as potential clients. Even so banks would consider family business as a client only it has injected itself into the formal economy, i.e. the banking system. When the family business moves from cash economy to a cashless and account-full economy, it graduates itself from the informal to formal sector of the economy. Having seen the trend of the money movement in the account, the bank may assess the family business' risk and extend its banking product and services according. The willingness of the bank to extend credit to the family business is also a form of investing in the family business. With the availability of capital, the family business can now move forward into the next stage of development: SMEs.

Indirect investment is general accomplished by a third part or entity, such as the government which brings benefits to the family business directly. These investments may not be in a form of money; it may be in a form of policy which leads to the increase productivity of the family business, such as tax exemption in certain types of business. The returns received by the government are the increase in revenue from other sources. For instance, if tax exemption for the family business leads to the reduction in cost, with that savings the family business passes on the savings to consumers in lower cost and thus stimulating consumption ---which in turn leads to higher GDP growth. The government in the process also collects additional revenue from tax collection. This type of investment is indirect and is not often seen by direct observation. The benefits of this type of investment generally appear one fiscal quarter or year later after the implementation of the policy.

5.6 Family Business as an Indicator of the Health of the Economy

Earlier discussion about the role of the family business as an indicator of the health of the economy focuses on the success of family business. There might be instances which the family business fails due to foreign investment in the country. For instance, the introduction of hypermarkets in Thailand resulted in keener competition against local retailers who are small. Larger hypermarkets, such as Tesco Lotus and Big C can compete on the basis of price and aim to secure market share through price cutting. Consumers rightfully flock to these hypermarkets and left the local mom-and-pop shops facing the cold reality of 'natural selection.' Although these local shops are family owners, they are neither the subject nor the focus of this book. These local shops are not competitive in the market; they represent a defect of pricing and imperfect competition. There threatened existence foretells a bigger story: economic evolution.

As these local shops fail, they open opportunities for larger firms to enter the market. Generally, these hypermarkets are foreign owned or foreign subsidiaries. In this case, it is antithetical to the earlier assertion that the failure of family business is an indicator of poor economic performance of the country. However, in this case, the failure of these "family businesses" signals opportunities for growth for foreign companies and increase utility for consumers who would otherwise have to deal with price hackling with local merchants. Therefore, this case demonstrates that failure of local businesses could make the country attractive to foreign investment.

5.7 Domestic Direct Investment (DDI) & Foreign Direct Investment (FDI)

The pooled fund used in the family business is a direct domestic investment (DDI). This type of investment often occurs in the informal sector: family business. As such, the amount of such an investment may escape public recordation. The actual amount invested in the family business sector may be larger than what may be reported.

DDI in the domestic economy is an evident of the self sufficiency of the country. The money invested in the family comes from private savings of family members. If the family business is successful and generates more income and further investment into the business, the informal sector may play a larger role in the economy than what we originally though. How big a role does the family play in the economy, we may not know for certain.

If DDI exists, the question then follows: *How can DDI be measured?* It can be measured through the tax system. In advanced economies where the tax system can capture business income and expenses through the income tax system, DDI may be determined with high level of precision. In fact, if the DDI rate could be compared to FDI rate, the number could forecast a new trend in the economy. However, research in family business has not yet progressed this far. Many researchers are still intoxicated with the lies of 'trans-generational transfer of ownership' and the decision making process of family business through the so-called 'family council.' These topics do their job well in filling up spaces in books and journals; they contribute nearly nothing to the advancement of knowledge in the field for lack of quantitative proof.

FDI is foreign direct investment. This investment depends on the health of the economy of the country. If the economy performs poorly, FDI inflow may be less. ***How can the health of the economy be measured?*** One means of that measurement is the GDP growth rate. However, the GDP is summarized every quarter at most and annually to be more accurate. Investment decision needs facts that are more expeditious. The GDP may be a poor indicator for foreign investors. The alternative may be DDI. However, the problem with DDI is that it is a new concept, a concept introduced by this book. If it could be compiled and published country-by-country, it would be a useful tool to supplement the decision making process for FDI.

5.8 Business Models and their Implications for Family Business

Family business is the outgrowth of an entrepreneurial enterprise of an enterprise. An entrepreneurial enterprise of an entrepreneur generally is a one man show. The entrepreneur takes risk in an enterprise by investing with known cost or expenses for the return of uncertain income. As time passes, the enterprise becomes successful, the entrepreneur involves other people within his circle. The closes associates of an entrepreneur are the family members. By incorporating family members into the enterprise, the entrepreneurial enterprise is transformed into a semi-structured business. In time, the family business may also grow into SMEs and finally reach the stage of becoming a corporation. This model of growth is a linear in form.

Can there be an alternative model where the family business starts out as a family business unit without going through the entrepreneurial path? There may be instances which the family business starts out as a 'family business.' In such case, the family entity generally is endowed with assets or liquid assets which could be employed in the enterprise. Having bypassed the natural growth process, the family business under this alternative model has greater risk exposure. *Risk* is defined as uncertainty of event occurrence that affects the business enterprise. Unlike the family business that had gone through the entrepreneurial growth as the first stage, the family business starting out as a family business unit at its inception has little knowledge or experience in the business; therefore, it has greater exposure to risk in the operation. In fact, many family business failed due to the lack of knowledge in the business. The availability of financial assets may allow the family business to start off with confidence and greater degree of involvement of family members; however, at the same time, the business is exposed to uncertainty.

Knowledge gained through experience by the entrepreneur renders the family business that followed a greater degree of leverage from the inventory of knowledge gained through years of experience. A family business that bypassed the natural path of business growth does not enjoy the same degree of leverage from experience and knowledge gained through prior operation. *Experience* is the practical knowledge gained through past operation which can be reused or shared within the organization for purposes of increase economic productivity and risk reduction. *Knowledge* is the inventory of 'know-how' retained by the agent through learning or experience that could be reused or stocked as inventory for future use. A family business that starts out at the second stage without going through the first lacks both experience and knowledge.

In case where the family business was acquired or transferred through succession in interest, the same rule applies. By inheriting the business or receiving the business through transfer, the successor in interest does not necessary also acquired the experience and knowledge of that business. Knowledge sharing and knowledge experience transfer may not take place unless the succession process is well prepared. Succession in interest as a model for family business is treated in the later paper.

6.0 The Theory of Firm

The theory of firm describes the firm's organization, structure, behavior and its relationship to the market (Kantarellis, 2007). In explaining how firms operate, the following theories had been used in the mainstream academic: (i) transactions theory, (ii) managerial and behavioral theory, (iii) Williamson's approach, and (iv) firm's economies. Other theories will also be explored in this section.

Transaction theory looks at the transaction cost. Firm arises as the result of cost differential among different source of production. When the cost of production outside of the firm is lower than if the firm produces the product for itself, there is a demand that a firm concentrating on producing that product. In words, through imperfect information and the imperfection of market, there will always be some firm who can produce a product to meet the needs of others. This creates a demand for assigned production to firms. Firms exist as the result of market imperfection. This imperfection allows some firm to produce at lower cost level and thus concentrate on that product and sell to though who cannot produce the product---as cheaply as the seller (Coarse, 1937).

Managerial and behavioral theory is the second approach to explain firm. In response to the challenged lodged against the transaction theory of William Coarse, other theorists, such as William Baumol (1959 and 1962) (Baumol, 1962), suggests that managers attempt to maximize their utility. This is not the same as maiximization of profit. The focus of managerial behavioral is on utility maximization, not profit maximization. The focus is on the individual managers who try to maximize personal gains, such as high salary, compensation package, and prestige. Under the behavioral approach to firm, in a situation of uncertainty the decision of firm is an outcome of compromising all interests within the firm (Cyert and james, 1963).

Williamson's approach to the firm theory claims that the existence of the firm depends on the asset. This is known as asset specificity. Each firm is endowed with a certain assets. These assets are specific for a certain production. Firms that do not own the asset must negotiate with firms that possess the assets (Williamson, 1975). Firms that possess the assets can produce, but they themselves do not have the order or needs for production. Through market imperfection and imperfect access to information, forms with information for the needs of certain products negotiate with firms who own the assets but do not have market knowledge. Under the state of imperfection, each party wants to be opportunistic in the deal, i.e. attempting to maximize gains for itself. However, there is a check-and-balance mechanism which keeps all parties at bay. Each party will not allow itself to be so opportunistic as to damage its own reputation and, thus, handicap its future dealings (Oliar *et al.*, 2008). Each party will optimize its gain with the bound of reason.

Firm's economies explain the objective of firms' attempt to return to scale. Through various combinations of capital and labor, the firm will maximize its utility and profit at a point where the average cost of output is lowest. The optimization may be achieved through economies of scale or economies of scope. *Economies of scale* is the reduction of per unit cost with increase in output. This per unit cost reduction may be achieved through learning and concentration. The second means to reduce cost is through economies of scope (Archibald, 1987[2008]; Panzar and Willig, 1981; Tirole, J., 1988). *Economies of scope* refers offering of multiple product lines. Instead of offering one product, the firm may offer a portfolio of products. In so doing, it may reduce the risk of not selling one product while making profits by satisfying the needs of buyers with other products in the portfolio.

In light of these theories, which one can best explain family business model? There is no single theory that can explain family business enterprise. The transaction theory can provides some explanation for the existence of the family business. The family business fulfills a certain needs in the market, i.e. it can provide some products to the market a cheaper cost and, therefore, selling at a lower price. There is a certain degree of pride involved in owning a business; this emotional element also draws on the manager's behavioral theory. A family business is very asset specific. Once a family is settles on a particular business, it tends to stay in that line of business because its assets are limited and specific for a certain productive output. This asset-specific model falls within the Williamson's approach to firm. As for economies of scale and economies of scope, the application may not be uniform; they may be applicable to some family businesses and to others they may not.

6.1 Profit Maximization vs. Value Creation

Profit maximization is the behavior of firms. It cannot be said that the family business does not want to maximize profit; however, profit maximization is not the overwhelming goal of the family business. Profit maximization in the firm is the overwhelming goal of a firm because of the fiduciary owed by the firm to the shareholder. In a family business, this fiduciary does not exist. Family business is more concerned with self sufficiency and survival than in optimizing its share value. For this reason, profit maximization is not the overwhelming goal of the family business.

Profit maximization is defined as the optimal level of owner's equity divided by the number of share in the firm. This issue is not common in the family business because by its nature the family business is co-owned by family members. Being a closely held entity, the family business is more concerned with the maximization of value. The goal of the family business is value creation. *Value creation* is the innovative use of the existing resources to create value. For instance, the family business may have an unused asset; value creation is to turn that idle asset into a productive asset thus making money for the business. Family business wants to optimize its value through value creation.

Value creation is not value added. Value additivity is the increasing in cost in the production process thereby leads to the increase in the final price to the consumer. If the sale is maintained at the same level, the value added would create additional profit. However, in reality value additivity generally leads to the loss of competitive advantage of the firm. By increasing price, the sales generally suffer.

6.2 Blue Fish Swimming in the Red Ocean

Recall that the family business generally grows out of the success of an entrepreneur and his enterprise. As an entrepreneurial enterprise, the family business seeks and exploits opportunity to optimize its value. Through value creation the family business sustain its business and carries on its operation. As an opportunity seeker, the family business may be called 'blue fish.' However, there is a limitation as to how much opportunity can the family business seek and exploit. Unlike the entrepreneurial who acts for and by himself, the family business is not structured and involved many people. Having additional stakeholders in the operation limits the freedom of the family business to act freely.

The market in which both the entrepreneur and the family business exist is one filled with competition. There is no business that can guarantee against assault from competition. All businesses face competition. Success invites competition. If the family business is successful, its success will attract other players to enter the market. For that reason the market is always a 'red ocean.' Some success may be short lived others may have a longer life span. A family business is akin to a 'blue fish swimming in the red ocean.'

In order to survive in the market, the family business must innovate its business in order to stay ahead of the competition. Innovation leads to value creation. That value creation is the seed for further success in family business. The threat of competition to family business may be keener than other formalized businesses because of its limited capital and expertise. Unlike the corporation or larger sized companies in which the availability of funds may be more flexible, family business must maintain profitable operation. Larger companies may sustain loss over a period of time and can still make a return to profitability. However, the family business cannot afford that type of loss over a long period of time. This inability to sustain loss comes from the fact that a family business is limited by size and resources. Many family businesses go out of business because they cannot sustain loss during the period of economic down turn.

6.3 Key Elements for Success: Size, Scope, Speed, and Strategy

Many writers point to the product mixed of 4Ps as the key for success in business. However, this old concept may not hold persuasion in today's business environment. What sets a successful business from a failed business is not 4Ps but something else. Failed business had product, price, promotion and place. If the 4ps explains success in marketing mixed, then companies with these 4Ps would not fail. However, all failed businesses had followed these 4Ps concept, but they still failed. An alternative explanation is needed. Corporate success may come from other sources: convenience, commodity, cost, communication, channel, consumer, circumstances (Shimizu, 2003, 2009; Solis, 2011). These 7Cs may also be applied to family business.

While other writers talked about 4Ps and 7C that contribute to the success of business, this book proposed threat factors that may contribute to the business failure: 4Ss which contributes to the competition and failure of family business. These 4S' include: *size, scope, speed,* and *strategy*. Family business is small in size. In today's economy, success comes through hypercompetition. Large size means the company can absorb shock from the system and sustain unprofitable operations longer than small-sized businesses. A family business is small and is at a disadvantage.

The second element that leads to competitive advantage is scope. Large corporation can engage in a variety of businesses. These businesses serve as a portfolio component to help spread the risk of operations. If one product line does not perform, other products will perform and make up for the losses. The family business cannot afford this luxury. Due to limited capital and experience, the family business is concentrative in a particular business line and is completely exposed to system shock in times of keen competition or economic down turn.

The third element of success for large corporation and a potential threat to a family business is speed. Despite the availability of technology and information, the family business still operates in the informal sector and is disconnected from the mainstream market. As part of the informal sector, the family business does not have the advantage of a larger corporation in exploring and exploiting market information. Unable to move at a quicker speed may work against the family business. In order to maintain success, the family business owner must keep abreast with market changes and plan according to avoid being left behind.

Lastly, the fourth element that allow large corporation to succeed and may threat family business is strategy. Strategy is defined as planning and matching the available resources to the set goals. Corporations have the resources to engage in short-term, medium-term and long-term planning. It has the resources to match with these plans. The planning allows the corporation to stay focus. The family business lacks the resources to engage in multiple ranges planning. Many family businesses lack long-term planning. Success comes from daily operations. Lacking strategic planning, the family business is always faced with the threat of competition from larger firm in the market who can afford to engage in strategic actions.

6.4 Effective Family Business Model

What is a successful business model for the family business? There does not appear to be a single model that can explain the success of a family business. The various theories of firms explored in this paper may all apply to family business at different stages and in different aspect of the family business.

The performance of the family business must be looked at in context. It is an outgrowth of the entrepreneurial enterprise. As such, the family business retains that entrepreneurial spirit in the owner of the business. As an entrepreneurial enterprise, the entrepreneur explores and exploits the system at the same time. As a family business, the family business owner has already known what contributes to its successful operation. By having additional stakeholders into the business, the family business seeks to be system protector. The family business limits itself to a certain scope of business; this scope limitation is akin to business concentration or focus. The family business succeeds with focus, but at the same time, it also faces a threat from external shock. Changes in the market may exert threats to the family business unit, i.e. changes in technology or demographics, in order to maintain profitable operation, the business must adapt accordingly.

Two elements may help explain the success of a family business: value creation and adaptability. The family business strives to create value. Value creation comes from setting idle resources to create economic benefits or increase economic productivity through the use of existing resources. By retaining the inventive spirit from the entrepreneurial stage of its development, the family business may strive well as a value creator. Secondly, the small size and limited stakeholders in the organization allows the family business to shift operations and adapt to the changing environment in order to maintain profitable operations.

6.5 Men & Women Family Business Owners as Creators of SMEs

Entrepreneurship is gender blind. Business operations should also be gender blind. Success is measured by the level of profit. In family business, gender does not determine success or failure. However, this is not to say that discriminatory practices among relevant players in the system, such as banks and other lending institutions may use gender as an indictor for decision making. Discrimination of any kind, gender or otherwise is unclean and not green.

Gender issues may be a research issue in entrepreneurship and family business. However, in reality success in the family business is determines by the performance of the business. The performance of the business depends on the resources inside the organization and the use of that resource to exploit the market. As a system, the market is gender neutral. As market participants, economic agents (men and women, more male than female) are gender discriminatory. To raise gender consciousness is a positive thing if it contributes to the betterment of the lot of all market participants. However, if gender issue is raised as an obstacle to the path of success of the family business, it has no place in our discussion. Such research may do more damage than conferring benefits to society.

6.6 Destructive Entrepreneurship & Family Business

All businesses are system exploiter. Businesses exist in order to make profit for its owners. However, certain businesses profit comes at the expense of other people. These types of businesses are called destructive entrepreneurship. Certain businesses exist in a parasitic form, i.e. it depends on the exploitation of other businesses. Copycat businesses in the developing countries, for instance, are good examples. These copycat companies engage in the illegal copies of protected products, i.e. products protected by copyrights, patents and trademarks. Although these copycat companies make large amount of profits, they are not seen as successful. In business success should come from the sweat of the brows; stealing and pilfering the rights of others is a destructive behavior. Such ill-gotten gains, no matter how large, remain a loot from destructive behavior.

A family business partially depends on goodwill for its success. It cannot afford to soil its good name and reputation through destructive behavior, such as illegal bootlegging of protected products. 'Business' of that type is an illegal activity, and cannot be classified as a family business or entrepreneurship. A "thief" does not qualify himself as a businessman just be the mere fact that he maintains an office and has a sign hanging at the front door saying "Open for Business." In order to quality to use the word business to describe its operation, the enterprise must govern itself by a certain code of conduct or business ethics. 'Thou shall not steal" is equally applicable to business owner as to any church goers (Exodus 20:1-21, Deuteronomy 5:1-23).

7.0 CONCLUSION

The theme of this part 2 of 4 of *'Entrepreneur and Entrepreneurship'* series is the role of the family business in the economy. Family business is a further step in development from the informal and unstructured operation of the entrepreneurial enterprise by the entrepreneur who engages in a one man show. An entrepreneur takes anything that comes his way. As an opportunist, an entrepreneur is a true system exploiter. However, the family business is more structure. The involvement or more stakeholders in the operations confine the family business to a particular boundary of operation. With a clearly defined scope of business, the family business may concentrate on a particular type of business and engages in value creation. This characteristic of a family business: value creation, sets it apart from the conventional firm which engages in profit maximization. Value creation looks to innovative spirit of the entrepreneur that is stilled ingrained in the family business owner. The inventive spirit of the entrepreneur in the family business always looks for means to turn idle assets to production assets. Value creation is achieved through turning a 'source' into resource. The circle of stakeholders in the family business is small. The fiduciary is closely regulated within a close circle of family members; this close nit organization sets it apart from a larger and more formally organized corporation.

The last section of this paper considers the issue of gender in business and concludes that sexism in any form, gender treatment or otherwise, is neither green nor clean in business. There are researchers who try different degrees of success among male and female business ownership. If there is a difference in the degree of success, it is not due to the difference in gender; the inequity may come from discrimination. Discriminatory practices among lending institution, for instance, may contribute to the different degree of success among male and female owned businesses. This difference in the degree of success does not come from the difference in gender; it comes from gender-based treatment.

REFERENCES

Antonelli, C. (2003). *The Economics of Innovation, New Technologies, and Structural Change*. London: Routledge. ISBN 0415406439; andJohnson, Bjorn (2008). "Cities, systems of innovation and economic development." Innovation: Management, Policy, and Practice 10 (2/3): 146–155.

Ahlstrom, D. (2010). "Innovation and Growth: How Business Contributes to Society." *Academy of Management Perspectives*, **24**(3): 11–24.

Archibald, G.C. (1987 [2008]). "firm, theory of the," *The New Palgrave: A Dictionary of Economics*, v. 2, p. 357.

Baumol, W. (1962). "The Theory of Expansion of the Firm," *A.E.R.*, **52**(5); pp. 1078-1087.

Coase, Ronald H. (1937). "The Nature of the Firm." *Economica*, **4**(16): 386–405.

Cyert, Richard; March, James (1963). *Behavioral Theory of the Firm*. Oxford: Blackwell. ISBN 978-0-631-17451-6.

Exodus 20:1-21, Deuteronomy 5:1-23, "Ten Commandments,'' New Bible Dictionary, Second Edition, Tyndale House, 1982 pp. 1174-1175.

Fornahl, D.; Broekel, T.; Boschma, R. (2011). "What drives patent performance of German biotech firms? The impact of R&D subsidies, knowledge networks and their location." *Papers in Regional Science*, **90**(2): 395–418.

Kantarelis, Demetri (2007). *Theories of the Firm*. Geneve: Inderscience. ISBN 0-907776-34-5.

La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 1999. "Corporate Ownership around the World." *Journal of Finance*, **54**(2): 471–517.

Loewen, Jacoline (2008). *Money Magnet: Attract Investors to Your Business*: John Wiley & Sons. ISBN 978-0-470-15575-2.

National Joint Committee for the Communication Needs of Persons With Severe Disabilities (1992). Guidelines for meeting the communication needs of persons with severe disabilities [Guidelines]. Available from www.asha.org/policy or www.asha.org/njc. Copyright © 1992, National Joint Committee for the ommunication Needs of Persons With Severe Disabilities. doi:10.1044/policy.GL1992-00201.

Oliar, Dotan; Sprigman, Christopher Jon (2008). "There's No Free Laugh (Anymore): The Emergence of Intellectual Property Norms and the Transformation of Stand-Up Comedy." *Virginal Law Review*, *94*(8): 1787–1867.

O'Sullivan, David (2002). "Framework for Managing Development in the Networked Organisations." *Journal of Computers in Industry*, **47**(1): 77–88.

Panzar, J.C. and Willig, R.D. (1981). "Economies of Scope." *American Economic Review*, **71**(2), p p. 268–272.

Peilei, F. (2011). "Innovation capacity and economic development: China and India." *Economic Change and Restructuring*, **44**(1/2): 49–73.

Shimizu, Koichi (2009). "Advertising Theory and Strategies," 16th ed., Souseisha Book Company. (Japanese).

Shimizu, Koichi (2003). "Symbiotic Marketing Strategy,"4th ed., Souseisha Book Company.(Japanese).

Solis, B. (2011). *Engage!: The Complete Guide for Brands and Businesses to Build, Cultivate, and Measure Success in the New Web*, John Wiley & Sons, Inc. pp.201-202.

Spulber, Daniel F. (2009). *The Theory of the Firm*, Cambridge.

Steil, B.; Victor, D. G.; Nelson, R. R. (2002). *Technological Innovation and Economics Performance. A Council of Foreign Relations Book*. Princeton University Press.

Tirole, J. (1988). *The Theory of Industrial Organization*. "The Theory of the Firm," pp. 18–20. MIT Press.

Williamson, Oliver E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: The Free Press.

Current Account Balance Analysis of ASEAN Countries

Chanoknath Sutanapong★

About the author

★Chanoknath Sutanapong is an independent researcher. She may be reached by email at: Chanoknath.sutanapong@gmail.com

ABSTRACT

The purpose of this paper is to provide a practical tool for binomial proportion in two sample studies. This binomial proportion of two sample analysis is a special case because it deals with discrete data. A discrete data is a categorical data where the category of interest is designated as 1 else 0. The data used in this paper comes from the IMF's annual report for World Economic Outlook. We extracted the current account balance as a percentage of the GDP as the unit of analysis. The countries of focus are the 10 countries in the ASEAN. The research question is "which country in the ASEAN effectively manage its economy on the basis of current account balance as a percentage of the GDP?" We defined positive current account balance as 1 and negative value as 0. The study period runs from 2005 to 2014. We found that no country in the ASEAN had significant positive current account balance

Keywords: current account, GDP, international trade management, ASEAN

CITATION:

Sutanapong, C. and Louangrath, P.I. (2015). "Current Account Balance Analysis of ASEAN Countries." *Inter. J. Res. Methodol. Soc. Sci.*, Vol., 1, No. 2: pp. 18-24. (Apr. – Jun. 2018).

1.0 INTRODUCTION

The objective of this paper is to provide a statistical tool to analyze discrete data through the use of two sample proportion. Two samples analysis is a common practice in the industry where one sample may be divided into two portion and both are analyzed. The comparison studies of these two samples may provide a clearer insight into the original sample and, hence, their origin or process from which the original sample came.

The two sample proportion introduced in this paper is used for discrete data. Discrete data are those that may be classified as categorical data. Given a data set that is not quantitative, the researcher identifies a class or category of interest in the set. The data element that meets the criterion is coded with 1 and those that do not meet the criteria are marked with 0. Thus, at the end, we have two categories: (1,0); this is called categorical data.

The basis for categorical data analysis begins with the Laplace definition of the probabilities of success (p) and failure (q). These p and q are defined as: $p = (s = 1) / (n + 2)$ where s is the sum of 1 and the (1,0) codification, and $q = 1 - p$. (Laplace, 1886; de Moivre, 1716[2000]; de la Vallée-

Poussin, 1907; Kenney and Keeping, 1951; Mirimanoff, 1930; Upensky, 1937). The statistical significance test for discrete data is given by the DeMoivre-Laplace Theorem:

$$Z = \frac{S_{x=1} - np}{\sqrt{npq}} \tag{1}$$

where $S_{x=1}$ is the sum of all observation coded as 1, n is the sample size, p and q are defined above (Walker, 1985).

To illustrate this concept, the data of current account balance as a percentage of the GDP of the 10 countries in ASEAN are used. The research question is "which country in ASEAN effectively and efficiently manages its international trade flow?" We defined (i) *effective management of international trade* as positive CA balance as percentage of the GDP and (ii) *efficient* international trade management as statistically significant counts of 1 in the (1,0) categorization in the study period. We examined 10 countries over a period of 5 years spanning from 2010 to 2014.

2.0 DATA

Secondary data of the current account balance as a percentage of the GDP for the ASEAN 10 countries were used for this study. Since the subject matter deals with macro-economic data, only secondary data may be used. Primary data is not practicable.

The sample size of 5 years is based on the recommendation under the Anderson-Darling test that adequate sample size may be achieved at 5 elements. In this paper, we used 5 years of current account as a percentage of the GDP in 10 ASEAN countries from 2010 to 2014. The data was obtained from the IMF's annual; report of the World Economic Outlook; this data is presented in Table 1.

Table 1. ASEAN countries current account as percentage of the GDP FY2010-2014

Country	2010	2011	2012	2013	2014
Brunei	36.60	34.71	29.84	20.88	30.73
Cambodia	(9.29)	(5.89)	(8.19)	(12.98)	(9.82)
Indonesia	0.70	0.19	(2.66)	(3.18)	(3.09)
Laos	(16.49)	(15.27)	(26.04)	(28.40)	(20.03)
Malaysia	10.08	10.89	5.17	3.48	4.39
Myanmar	(1.10)	(1.84)	(4.00)	(4.88)	(2.17)
Philippines	3.60	2.52	2.78	4.19	3.78
Singapore	23.44	22.15	17.00	16.52	18.68
Thailand	3.37	2.54	(0.43)	(1.16)	3.74
Vietnam	(3.79)	0.17	5.96	4.54	4.88

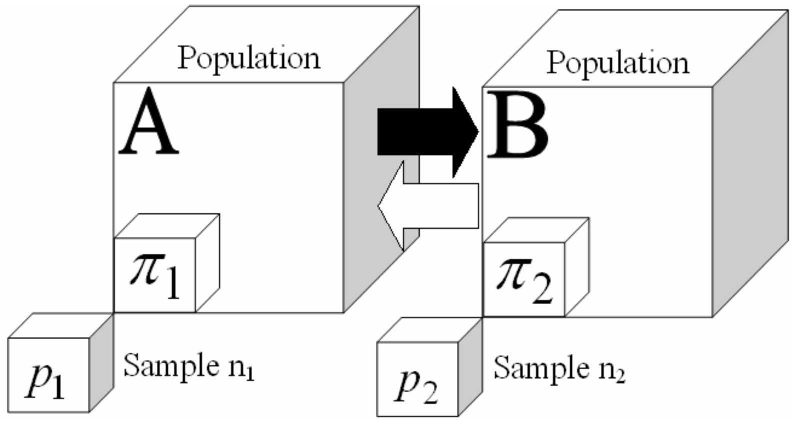
Source: https://www.imf.org/external/pubs/ft/weo/2018/01/weodata/index.aspx

The value in Table 1 is categorized into (1,0) by the following definition: positive current account balance as a percentage of the GDP is classified as success and is coded as 1; negative value is coded as 0. This process of discretization of the data is presented Table 2. After coding, the probabilities of success and failure with their attendant significance test are accomplished as defined in equation (1), *supra*.

3.0 METHODOLOGY

The objective is to study two sample proportions from two populations and make conclusion about the population. The case involves binomial distribution. There are two populations: populations A and B. In each population there are classes of interest: A and B. A sample is randomly selected from each population. Each sample has a sample size of n_1 and n_2 with a corresponding sample proportions of p_1 and p_2 respectively. The population proportion are given by π_1 and π_2. The structure of this comparison study is illustrated by the figure below.

Fig. Population proportion



The objective of the test is to determine whether the assumed equality of population proportion $\pi_1 = \pi_2$ based on two samples: one sample taken from each population. The null hypothesis claims that $\pi_1 = \pi_2$. The test compares the sample proportion and determine whether $p_1 = p_2$. If the samples are significantly different, then it is conclude that the population proportions are also different.

The two samples in this case are the two groups of the ASEAN countries. There are 10 countries in ASEAN. The 10 countries are divided as group 1 which of 6 countries: Brunei, Indonesia, Malaysia, Philippines, Singapore, and Thailand. Group 2 consists of Cambodia, Laos, Myanmar and Vietnam. The object of the test is to verify whether the two groups are significantly different. Under the ASEAN Economic Community (AEC) framework, the ASEAN community should be homogenous, at least in economic terms. Thus, as a test for this policy and vision of "oneness", we are using the current account balance as a percentage of the GDP as a proxy to test for their homogeneity. We opted for the current account balance as a percentage of the GDP because among the 10 countries, the GDP values are diverse. Singapore, for instance, show the highest level of GDP and Myanmar the lowest. However, this difference unbalanced diversity is controlled when we look at the current account balance as a percentage of the GDP; thus, selection bias is minimized.

3.1 Binomial Distribution

Binomial distribution is defined as random variable X with parameter n and p where X is the cumulative outcome, n is number of trials and p is the number of success. This binomial distribution may be written as $X \approx B(n, p)$ with a *probability mass function* given by:

$$f(k; n, p) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{2}$$

For $k = 0,1,2,\ldots,n$ and where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ reads "*n* choose *k*." The *n* is the number of trials and k is number of success (Papoulis *et al.*, 2002; Feller, 1968). This means *k* success and *n* - *k* failure. The probability of success is p^k. Failure is represented by $n-k$; therefore, the probability of failure is $(1-p)^{n-k}$. Recall that binomial distribution occurs in a two category answers: {yes | no} or {success | failure} outcome.

A random experiment involving two possible outcome: success or failure, in a repetitious fashion is called *Bernoulli Trials*. An illustrative example is the tossing of a coin. There are only two outcomes: either the coin would land head or tail. Successive repetition of the tossing would allow us to derive a rule of probability of the two possible outcome if the total number of tossing n and the number of *heads* turning up. Assume that "head" is classified as a "success" and success is symbolized by k. The probability of success may be written as p and the probability of failure (tail) is simply $1-p$. The distribution of the probability of the outcome is called *binomial distribution*. The preceding description is for Bernoulli distribution because it involves *one kind of Bernoulli trial*, i.e. coin tossing. If the trial involves many kinds of trials, i.e. coin, {yes | no} questions, at the same time, then a new type of binomial distribution is needed to explain the distribution: Poisson distribution (Poisson, 1837; p.206)

A Poisson sampling exists when in a given population, each element of the population is subject to a Bernoulli trial. Each element in the sample may have a different probability of success in the trials. The *probability mass function* for the Poisson distribution is different from that of the Bernoulli type:

$$f(k;\lambda) = P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{3}$$

where $k!$ = number of success (k factorial); e = Euler constant or base of the natural logarithm i.e. $e = 2.718...$; and λ = variance of the outcome: $\lambda = E(X) = Var(X)$.

The mean of the Poisson distribution is lambda: λ. The mean deviation is given by (Johnson *et al.*, 1993):

$$E\left|X-\lambda\right| = 2\exp(-\lambda)\frac{\lambda^{\lfloor \lambda \rfloor + 1}}{\lfloor \lambda \rfloor !} \tag{4}$$

The bound for the median (Choi, 1994) is given by:

$$\lambda - \ln 2 \le v < \lambda + \frac{1}{3} \tag{5}$$

The bound for the tails probability (Massimo *et al.*, 2007) is given by:

$$P(X \ge x) \le \frac{e^{-\lambda}(e\lambda)^x}{x^x} \quad \text{, for } x > \lambda \tag{6a}$$

$$P(X \ge x) \le \frac{e^{-\lambda^{x^x}}(e\lambda)^x}{x^x} \quad \text{, for } x < \lambda \tag{6b}$$

Therefore, the confidence interval for the Poisson distribution is:

$$\frac{e^{-\lambda^{x^{x}}}(e\lambda)^{x}}{x^{x}} \leq \lambda \leq \frac{e^{-\lambda}(e\lambda)^{x}}{x^{x}} \qquad (7)$$

The Poisson distribution comes from the *Poisson process*. Poisson process is a stochastic process which (i) counts the number of event, (ii) events occur within a fixed time interval: $\{N(t); t \geq 0\}$. For instance, telephone calls (Willkomm *et al.*, 2009) within a time interval (one hour) in a Customer Service Center. We will leave the subject of Poisson distribution at this introduction to different types of binomial distribution and will revisit the subject in future paper.

3.2 Test Statistic for the Z-test for equality between proportions in binomial distribution

Given two populations with a specified class of interest proportions of π_1 and π_2. Two random samples are drawn from each population. The proportions in the sample are given by p_1 for sample one corresponding to the population proportion π_1, and p_2 corresponding to the population proportion π_2. The test statistic to determine the level of significance of the difference between the population proportions on the basis of sample proportion, i.e. determine the sample proportion and make a conclusion about the population. The test statistic for the equality between two proportions in binomial distribution is given by:

$$Z = \frac{(p_1 - p_2)}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (8)$$

where $P = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$.

The hypothesis formulated as: $H_0 : \pi_1 = \pi_2$ and $H_A : \pi_1 \neq \pi_2$. Reject the null hypothesis if $\pi_1 \neq \pi_2$. The test may be one-tailed or two-tailed. The Z-critical is given by $Z_{\alpha}(0.05) = \pm 1.96$.

4.0 FINDING AND DISCUSSION

We present the findings in two parts. First, we treated the data under continuous distribution and verify the significance level of the countries that had effective current account balance as a percentage of its GDP. We define effective current account balance as a positive value. Negative account balance comes from excessive imports; we classified such a case as ineffective account balance management. Under this approach to evaluate the country on macro-economic data, the GDP alone is not adequate. By its nature unless the country suffers from economic crisis of shock, the GDP general grows from year to year. However, the current account balance status tells us whether that growth in the GDP stays in the country or does it grow numerically, but at the same time flow out of the country to pay for imports. Thus, despite GDP growth, if the current account balance is negative, the GDP growth is not meaningful.

4.1 Significance test in continuous probability

Under continuous probability analysis, we found that Brunei has a significant CA as percentage of the GDP in the five years period that we studied ($p = 0.02$). As a group, the ASEAN has effective international trade management as measured by CA as percentage of the GDP with exception of one year in 2013 where the number was negative; the average for the five years period (2010 – 2014) is 4.57 ± 2.67. In 2006, there was a significant improvement in CA for the group ($p = 0.04$).

Table 3. Significance level for current account balance by country and group under continuous probability analysis

Country	Per Country	ASEAN	Z	F(Z)	*p* = 1 – F(Z)
Brunei	35.94*	6.25**	2.00***	0.9782	0.02
Cambodia	(6.90)	9.19	(1.08)	0.1392	0.86
Indonesia	(0.18)	7.49	(0.52)	0.3029	0.70
Laos	(17.51)	3.75	(1.43)	0.0756	0.92
Malaysia	11.00	4.36	0.45	0.6724	0.33
Myanmar	(0.71)	4.71	(0.36)	0.3577	0.64
Philippines	3.50	5.02	(0.10)	0.4593	0.54
Singapore	20.24	1.94	1.23	0.8913	0.11
Thailand	1.92	(0.10)	0.14	0.5540	0.45
Vietnam	(1.60)	3.11	(0.32)	0.3757	0.62

*Country value is the average over a period of 5 years: 2010 – 2014. **ASEAN value is the ASEAN average over a period of 5 years: 2010 – 2014. ***Standard score: $Z = (X_i - \bar{X}) / S$ where Xi is for country's average, $\bar{X}$ is the ASEAN mean and *S* is the ASEAN standard deviation.

4.2 Significance test in discrete probability

The results under discrete probability analysis are similar to that found under continuous probability method. For individual countries, there is no member country in ASEAN that show significant positive CA as percentage of the GDP ($p > 0.05$). As a group, the average probability for success in having positive CA as percentage of the GDP is 0.53 ± 0.04.

Table 4. Significance level for current account balance by country and group under discrete probability analysis

Country	2010	2011	2012	2013	2014	pValue
Brunei	1	1	1	1	1	0.68
Cambodia	0	0	0	0	0	0.82
Indonesia	1	1	0	0	0	0.47
Laos	0	0	0	0	0	0.82
Malaysia	1	1	1	1	1	0.68
Myanmar	0	0	0	0	0	0.82
Philippines	1	1	1	1	1	0.68
Singapore	1	1	1	1	1	0.68
Thailand	1	1	0	0	1	0.34
Vietnam	0	1	1	1	1	0.32
p ASEAN	0.58	0.67	0.50	0.50	0.58	
q	0.42	0.33	0.50	0.50	0.42	
Z	0.11	0.22	-	-	0.11	
F(Z)	0.54	0.59	0.50	0.50	0.54	
pValue	0.46	0.41	0.50	0.50	0.46	

The pValue is defined as *1 – F(Z)* and *F(Z)* is defined as:

$$F(Z) = \frac{1}{1 + \exp(-\sqrt{\pi}\left(\beta_1 Z^5 + \beta_2 Z^3 + \beta_3 Z\right)} \qquad (9)$$

where $Z = (X_i - \bar{X}) / S$, β_1 = 0.0004406, β_2 = 0.0418198, and β_3 =0.90000000.

5.0 CONCLUSION

The aim of this paper is to answer the question of how to measure effective international trade management? We used the current account balance as percentage of the GDP as the indicator. Ten countries of the ASEAN were used as proxy. We analyze the ASEAN countries as a group and individual countries. The study period span from 2010 to 2014. We classified this significance as *efficient* international trade management, and effective management of international trade is defined as positive CA as percentage of the GDP. No country in ASEAN qualified as *efficient* in international trade management under 95% confidence interval. Among the 10 countries, Thailand and Vietnam scored the highest at 0.6620 and 0.6728 probability of success (effective management of international trade) with pValues of 0.34 and 0.32 respectively which exceeds the allowable significance level of $p < 0.05$. As a group, ASEAN or the AEC (ASEAN Economic Community) has and average of 0.53 level of effectiveness in international trade management ($p = 0.47$).

REFERENCES

Choi K.P. (1994). "On the medians of Gamma distributions and an equation of Ramanujan." *Proc. Amer. Math Soc.*, **121**(1) 245–251.

de la Vallée-Poussin, C. (1907). "Demonstration nouvelle du théorème de Bernoulli." *Ann. Soc. Sci. Bruxelles*, **31**, 219-236.

de Moivre, A. (1716). *The Doctrine of Chances, or, a Method of Calculating the Probabilities of Events in Play*, 3rd ed. New York: Chelsea, 2000. Reprint of 1756 3rd ed. Original ed. published 1716.

Feller, W. (1968). An Introduction to Probability Theory and Its Applications. Volume 1. Wiley. Section VII.3. ISBN 0-471-25708-7.

Johnson, N.L., Kotz, S., Kemp, A.W. (1993). *Univariate Discrete distributions*, (2nd edition). Wiley. ISBN 0-471-54897-9, p157.

Kenney, J. F. and Keeping, E. S. (1951). "The DeMoivre-Laplace Theorem" and "Simple Sampling of Attributes." §2.10 and 2.11 in *Mathematics of Statistics*, Pt. 2, 2nd ed. Princeton, NJ: Van Nostrand, pp. 36-41.

Laplace, P. (1886). *Théorie analytiques de probabilités*, 3ème éd., revue et augmentée par l'auteur. Paris: Courcier, 1820. Reprinted in *Œuvres complètes de Laplace*, tome 7. Paris: Gauthier-Villars, pp. 280-285.

Massimo Franceschetti and Olivier Dousse and David N. C. Tse and Patrick Thiran (2007). "Closing the Gap in the Capacity of Wireless Networks Via Percolation Theory." *IEEE Transactions on Information Theory*, **53**(3): 1009–1018.

Mirimanoff, D. (1930). "Le jeu de pile ou face et les formules de Laplace et de J. Eggenberger." *Commentarii Mathematici Helvetici*, **2**, 133-168, 1930.

Poisson, S.D. (1837). Probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilitiés (Paris, France: Bachelier, 1837), p. 206.
https://books.google.co.th/books?id=uovoFE3gt2EC&pg=PA206&redir_esc=y#v=onepage&q&f=false

Papoulis, Athanasios; Pillai, S. Unnikrishna (2002). Probability, Random Variables, and Stochastic Processes (4th ed.). Boston: McGraw-Hill. ISBN 0-07-122661-3.

Uspensky, J. V. (1937). "Approximate Evaluation of Probabilities in Bernoullian Case." Ch. 7 in *Introduction to Mathematical Probability*. New York: McGraw-Hill, pp. 119-138.

Walker, Helen M (1985). "De Moivre on the law of normal probability." In Smith, David Eugene. A source book in mathematics. Dover. p. 78. ISBN 0-486-64690-4. https://www.york.ac.uk/depts/maths/histstat/demoivre.pdf

Willkomm, D.; Machiraju, S.; Bolot, J.; Wolisz, A. (2009). "Primary user behavior in cellular networks and implications for dynamic spectrum access." *IEEE Communications Magazine*, **47**(3): 88.

New Log Likelihood Estimation Function

Louangrath, P.I. ★

About the author

★Louangrath, P.I. is an Assistant Professor in Business Administration at Bangkok University, Bangkok, Thailand. He could be reached by email at: Lecturepedia@gmail.com

ABSTRACT

This paper provides a New Log-Likelihood Estimator (NLLE) function as a tool for value approximation. We improved the accuracy of the log MLE in two steps (i) determine the log likelihood of a random variable *X*, and (ii) adjust the estimate by a factor of $(1/n-1)(\ln L(X))$. In-Sample testing was accomplished by using daily SET100 indices over a period of 60 days. Out-of-sample data were used for confirmatory verification; out-of-sample data came from 5 major stock markets: NASDAQ, DOW, SP500, DAX, and CAC40. Relevant tests used to compare the results of the proposed NLLE include Cramer-Rao Lower Bound (CRLB), Likelihood Ratio Test, Wald statistic, and Lagrange Multiplier (Score Statistic). It was found that NLLE is more efficient than the conventional MLE. It gives practitioners a better tool for value estimation in many fields of natural and social sciences.

Keywords

Cramer-Rao Lower Bound (CRLB), maximum likelihood estimator (MLE), Monte Carlo, Lagrange multiplier (Score Statistic), likelihood ratio test, log likelihood estimator (LLE), and Wald statistic.

JEL CODE: C10, C13, C14, C46, E27. G11, G17

CITATION:

Louangrath, P.I. (2015). "New Log Likelihood Estimation Function." *Inter. J. Res. Methodol. Soc. Sci.*, Vol., 1, No. 2: pp. 25-35. (Apr. – Jun. 2018).

1. INTRODUCTION

The research question presented in this paper is: "whether the current log-likelihood function is adequately accurate for value estimation?" Researchers always seek better tool for stock price forecast or value estimation. The likelihood function is a common tool for value estimation (Myung, 2003). The current likelihood function lacks precision (Efron, 1981). This paper addresses this weakness in two steps: (i) determine the log likelihood of $(X_i : x_1,...,x_n)$, and (ii) adjust the estimate by a factor of $(n-1)(\ln L(X))$. The data used for this research consists of 5 major stock markets: NASDAQ, DOW, SP500, CAC40, FTSE. over a period of 60 days. The minimum sample size requirement was met by using the *n*-omega method (Louangrath, 2014).

There is a gap in the literature concerning the improvement of likelihood function. For example, there is the need for an accurate forecasting tool to estimate stock price in face of price fluctuation. Under price fluctuation, the current MLE method cannot provide accurate estimation of the expected value. A more accurate estimation tool would have practical application in investment risk management. This paper attempts to fill that gap. In this paper, $\{X_i\}$ represents stock price. Two general cases classified by data types introduce the subject matter of likelihood function and the log-likelihood function as tools for estimating the expected value for random variable $\{X_i\}$.

This paper is presented in five sections. Section 1 introduces the likelihood function and its corresponding log-likelihood as the motivation for stock price analysis in light of price fluctuation. Section 2 explains the data used for the Monte Carlo simulations. Section 3 presents the proposed correction to the log-likelihood estimator and introduces a predictive function to track and forecast stock price under volatile condition. The findings and discussions are covered in section 4. Section 5 concludes the paper by urging practitioners to employ the new LLE as a new tool for stock price estimation and risk management.

2. DATA

The data used for the simulation and testing in this research came from two sources. The first set of data is comprised of the daily CLOSE price of the 100 companies comprising SET100 index for the Stock Exchange of Thailand. The daily prices of these 100 over a period of 30 training sessions were used as a sample. A second set of data consists of the daily indices of 10 major stock markets: (i) NASDAQ; (ii) DOW; (iii) SP500; (iv) CAC40; and (v) FTSE. This second set of data is used for out-of-sample confirmatory verification so that the proposed LLE could be generalized. The generalizability of the proposed LLE is based on the result of consistency in Monte Carlo simulation of 217 iterations using out-of-sample data.

3. METHODOLOGY

The validity and reliability of the old and new log-likelihood functions are compared. The in-sample employed the data from the Stock Exchange of Thailand. Out-of-sample tests were achieved through the use of stock indices taken from 10 major stock markets over a period of at least 60 trading sessions. Comparative results of the following tests were observed: Cramer-Rao Lower Bound (CRLB), Likelihood Ratio Test, Wald statistic, and Lagrange Multiplier (Score Statistic).

The log-likelihood function in the current literature may be written linearly as:

$$\ln L(\mathrm{X}) = \frac{1}{n}\sum\left(\mathrm{X}_i \ln L(F(Z) + (1-\mathrm{X}_i)\ln L(1-F(Z))\right) \quad (1)$$

where X is the observed values $(\mathrm{X}_i : \mathrm{x}_1,...,\mathrm{x}_n)$, Z is the standard score given by $Z = (X_i - \overline{X})/S$ and $F(Z)$ is the percentage probability read from the Z-table at a given critical Z value. In this study, X_i is the CLOSE price of the stock in the SET100 index. For the out of sample test, X_i represents the daily market indices of 5 major stock markets. The result of (1) is called the expected value of series $(\mathrm{X}_i : \mathrm{x}_1,...,\mathrm{x}_n)$ or $E[\mathrm{X}]$. The result of the estimate is compared to the arithmetic means. The arithmetic mean is given by: $\overline{X} = (1/n)\sum X_i$.

The arithmetic mean is a biased estimator if the difference between the arithmetic mean and the expected value is non-zero: $\overline{X} - E[X] \neq 0$. In the learning data of 15 days of trading, the arithmetic mean of the PTT per share price is $\overline{X} = 237.00$, but the estimated value under the log likelihood method is 549.39 Baht per share. The over-shoot is twice the arithmetic mean. The current MLE method does not produce adequately precise value for small sample.

This paper proposes to modify the log likelihood function in order to reduce the bias or inaccuracy between the arithmetic mean and the expected value. The proposed function consists of two steps: (i) determine the log likelihood of $(\mathrm{X}_i : \mathrm{x}_1,...,\mathrm{x}_n)$ under the new LLE, and (ii) adjust the estimate the LLE in (i) by subtracting $\left(\ln L(X)\right)^{-n-1}$. The proposed log-likelihood in step 1 is given by:

$$\ln L(X) = \sum\left[\left(\frac{X_i \ln F(Z)}{n-1}\right)+\left(\frac{(1-X_i)(1-F(Z))}{n^2-n-1}\right)\right] \tag{2}$$

With the estimate obtained in (2), the expected value for $(\mathrm{X}_i : \mathrm{x}_1,...,\mathrm{x}_n)$ is obtained through:

$$E[X] = \frac{1}{n}\sum\left[\left(X_i - \ln(X_i)\right) - \left|\frac{\sum \ln L(X)}{n-1}\right|\right] \tag{3}$$

The result of (3) is approximates equal to $\bar{X} \cong E[X]$ and, thus, the bias for the estimator is minimized. Under this proposed log-likelihood estimator function, the arithmetic mean approximately equals to the expected value. Note that if the denominator of the correcting factor is n such that:

$$\frac{1}{n}\sum\left[\left(X_i - \ln(X_i)\right) - \left|\frac{\sum \ln L(X)}{n}\right|\right] \tag{4}$$

The expected value is observationally equivalent to the arithmetic mean of the series $(\mathrm{X}_i : \mathrm{x}_1,...,\mathrm{x}_n)$, this breaks the first property of *identity* where the condition $\hat{\theta} \to \theta$, but $\hat{\theta} \neq \theta$ must be maintained. Therefore, in (3) following the Bessel correction, the degree of freedom is used. Under the new LLE method, the Fisher information ($I(\theta_0)$) is calculated by $I(\theta_0) = 1/\operatorname{var}[X_i - \ln L(X_i) - |\ln L(X_i)/n-1|]$.

3.1 Monte Carlo Simulation

Monte Carlo is a class of algorithm used in repeated measurement as a simulation tool to approximate the true value (Kroese et al., 2014). One function served by the Monte Carlo simulation is to iterate the measurement until asymptotic normality is achieved or series of estimated values is stabilized. The issue of Fisher information as a means for explaining asymptotic normality is a common approach in the literature. Although theoretically sound, this approach may not be practicable in a case of smaller sample size and data volatility. In the context of stock price movement, investors are expected to see price volatility within a short period. For purposes of risk management, decisions must be made shorter time frame. Therefore, the Monte Carlo iteration size must be adjusted accordingly.

Define that $\hat{\theta} = T$ and $\theta = Z$ then the asymptotic normality condition $\sqrt{n}\left(\hat{\theta}-\theta\right) \to N(0,\sigma^2)$ becomes $\sqrt{n}\left(T-Z\right) \to N(0,\sigma^2)$. Since the sample and population distribution under normality is almost always equal, their relationship may be written as $\sigma^2 = S^2$. Therefore, the argument $\sqrt{n}\left(T-Z\right) \to N(0,\sigma^2)$ may also be written as

$\sqrt{n}\left(S^2 - \sigma^2\right) \to N(0, \sigma^2)$. Under this approach, the variance of the two distributions is used as the basis for the analysis. In order to test that this proposed method to assure that it is reliable, we suggest the use of repeated measurement through Monte Carlo simulation. Monte Carlo is given by:

$$\lim_{n \to R} \Pr\left[\frac{1}{N}\sum_{i=1}^{N} \xi - \mu\right] \le \frac{3\sigma}{\sqrt{n}} = 99.80\% \tag{5}$$

Given a series of random variable $\{X_i\}$, there exists the maximum and minimum values, the error for Monte Carlo is given by:

$$\xi^* = \left(\frac{\max - \min}{2}\right)/50 \tag{6}$$

The number of simulation needed in Monte Carlo is obtained through:

$$N = \left(\frac{3\sigma_{xi}}{\varepsilon^*}\right)^2 \tag{7}$$

where σ_{xi} is the estimated standard deviation of three values: $X_1 = \max$, $X_1 = \min$, and $X_3 = (\max + \min)/2$. Similarly, μ in (24) $\mu = \bar{X} - \left(T\left(S/\sqrt{n}\right)\right)$ The learning data of PTT price for 15 days produces the following N iteration for the series. In the initial example of 15 days of PTT price, the maxima is 250 and the minima is 209. Thus, $X_1 = 250$; $X_1 = 209$; and $X_3 = (250 + 209)/2 = 459/2 = 229.5$. Use these three X's to determine μ. The standard deviation of the 3 X's is $S = 20.15$ with the corresponding estimated standard deviation of $\sigma_{xi} = 16.74$. For this information, under (7), the expected error is $\xi^* = [(250 - 209)/2]/50 = 0.41$. The Monte Carlo iterations size under (7) is simply: $N = ([3(16.74)]/0.41)^2 = 122.49^2 = 15{,}003$. This means that with a sample of 15 days trading, the Monte Carlo requires 15,003 iterations in order to obtain an acceptable estimated value of stock price.

In the case of 5 major stock market indices out of sample test, the required Monte Carlo iteration is 22,500 runs. The required Monte Carlo iteration for SET100 components is 217 using adjusted result from the new LLE. For practical purpose, this conventional approach may not be practical. In stock trading where decisions are required within shorter span of time and the available of sample size is also smaller, demands for efficiency requires a faster method.

The new LLE method approximates the value of the set as $E[X]$ and compared it to the arithmetic mean of the sample ($\bar{X}$). These two values are used as the maximum and minimum values to obtain the Monte Carlo iteration. The required number of iteration is reduced considerably. This reduction evidences the efficiency of the new LLE function.

4. FINDINGS & DISCUSSION

The proposed new log-likelihood estimation (LLE) method provides a better estimation for the expected value of a given series of random variable $(\mathrm{X}_i : \mathrm{x}_1, \ldots, \mathrm{x}_n)$. The improvement is evidenced through the result of the following tests: Cramer-Rao Lower Bound (CRLB), Likelihood Ratio Test, Wald statistic, and Lagrange Multiplier (Score Statistic). This improvement was achieved without sacrificing the general requirements of MLE: (i) consistency, (ii) asymptotic normality, and (iii) efficiency.

For the SET100 data set, the general finding is that the proposed new log-likelihood function produces more efficient estimation while retaining acceptable level of Fisher information. When subjected to hypothesis tests, the new LLE performed well.

4.1 Cramer-Rao Lower Bound Test

The Cramer-Rao Low Bound test is used to verify the efficiency of the proposed new LLE method. Efficiency is defined as the optimality of the estimator, i.e. experimental design (Everitt, 2002) or hypothesis testing procedure (Nikulin, 2001). More efficient procedure needs less observations, i.e. if the model is efficient, the required sample size is smaller. The efficiency of an unbiased estimator, T, for parameter θ is defined as:

$$e(T) = \frac{1/I(\theta)}{\text{var}(T)} \qquad (8)$$

where $I(\theta)$ is the Fisher information of the sample and $e(T)$ is the minimum possible variance of an unbiased estimator divide by its actual variance (Fisher, 1921). The Cramer-Rao bound is used to prove that $e(T) \le 1$. Efficiency is achieved at $e(T) = 1$. This is proved by the Cramer-Rao inequality for θ. The Cramer-Rao bound is given by:

$$\text{var}(\hat{\theta}) \ge \frac{1}{\text{I}(\theta)} = \frac{1}{-E\left[\frac{\partial^2}{\partial\theta^2}\log f(X \mid \theta)\right]} \qquad (9)$$

Table 1. Comparison of Results from Old and New LLE

Market	$\overline{X}$	Old LLE	Diff*	New	Diff**
Dow	17,614.72	13,548.25	-4,066.47	17,708.73	94.01
SP500	2,080.28	1,382.22	-698.06	2,095.27	14.99
NASDAQ	5,036.54	4,473.78	-562.76	5,037.91	1.37
DAX	4,927.64	3,245.74	-1,681.90	4,896.49	- 31.15
CAC40	11,117.15	12,470.79	1,353.64	11,103.28	- 13.87
TWSE	8,476.10	9,847.43	1,371.33	8,562.26	86.16
Heng Seng	24,097.87	27,551.49	3,453.62	24,303.04	205.17
Shanghai	4,032.44	5,389.37	1,356.93	3,693.45	- 338.99
KOSPI	2,025.44	2,063.35	37.91	2,020.38	- 5.06
NIKKEI	20,240.28	- 1,371.32	-21,611.60	20,403.06	162.78

* $\overline{X}$ - Old LLE = Off from the arithmetic mean. ** $\overline{X}$ - New LLE = Off from the arithmetic mean.

The result of the Cramer-Rao test is shown in Table 2. The efficiency achieved under the new log-likelihood function is at or near 1.00. This is consistent with the findings in Table 1 where the accuracy of the old LLE is about 80% and the accuracy of the new LLE is 99.82% which also meets the requirement of the Monte Carlo simulation for $3\sigma/\sqrt{n} = 99.80\%$.

Table 2. Result of the Cramer-Rao Test under Conventional MLE and New LLE

Markets	$Var(\hat{\theta})$	$I(\theta)$	$eT = \frac{1/I(\theta)}{\text{var}(T)}$	$Var(\hat{\theta}) \ge \frac{1}{I(\theta)}$
Dow	308,045.00	0.00000	1.00	Yes
SP500	4,505.00	0.00022	1.00	Yes
NASDAQ	11,268.00	0.00009	1.00	Yes
DAX	51,553.00	0.00002	1.00	Yes
CAC40	15,654.00	0.00006	1.00	Yes

4.2 Likelihood Ratio Test

The likelihood ratio test is based on chi square distribution with degree of freedom of $df = df_2 - df_1$ (Huelbeck, 1997). The ratio calculation is the likelihood of the null divided by the likelihood of the proposed model. The test statistic was given by as $\Lambda(x)$ by Wilk (1938) as:

$$\Lambda(x) = \frac{L(\theta_0 \mid X)}{L(\theta_1 \mid X)} \tag{10}$$

or equivalently:

$$\Lambda(x) = \frac{L(\theta_0 \mid X)}{\sup\{L(\theta \mid X) : \theta \in \{\theta_0 m \theta_1\}\}} \tag{11}$$

where $L(\theta \mid X)$ is likelihood function, *sup* is the supremium function. The decision rule is governed by if $\Lambda > c$ do not reject the null hypothesis and if $\Lambda < c$ then reject the null hypothesis. The rejection point is the probability $\Lambda = c$. The variable c and q are selected at specified alpha (error) level whose relationship may be summarized as: $qP(\Lambda = c \mid H_0) + P(\Lambda < c \mid H_0) = \alpha$. The likelihood ratio test is a tool against Type I error. Type I error occurs when the null hypothesis is wrongly rejected. In the seminal literature, the likelihood ratio test has been classified as a power test (Neyman & Pearson, 1933). Casella and Berger (2011) wrote (10) and (11) as:

$$\Lambda(x) = \frac{\sup\{L(\theta \mid x) : \theta \in \theta_0\}}{\sup\{L(\theta \mid x) : \theta \in \theta\}} \tag{12}$$

Equations (10), (11) and (12) yield the same result.

The calculation for the likelihood ration follows a chi square hypothesis testing. With 60 counts for each market, the ratio is 1 or near one for all markets. This near 1 result shows that the estimation is close to the actual observed value or arithmetic mean. The critical value against which the ratio is test is 79.10. The null hypothesis that the two groups are not significantly different cannot be rejected.

Table 3. Result of the Likelihood Ratio Test under Conventional MLE and New LLE

Market	$L(\theta_0 \mid X)$	$L(\theta_1 \mid X)$	$\Lambda(x)$	$\chi^2(60) = 79.10$
Dow	17,614.72	17,708.73	0.99	Not significant
SP500	2,080.28	2,095.27	0.99	Not significant
NASDAQ	5,036.54	5,037.91	1.00	Not significant
DAX	4,927.64	4,896.49	1.01	Not significant
CAC40	11,117.15	11,103.28	1.00	Not significant

4.3 Wald Statistic

The third test to assess the likelihood function is the Wald statistic. For a single-parameter scenario, the Wald statistic is given by:

$$W = \frac{\left(\hat{\theta} - \theta_0\right)^2}{\operatorname{var}\left(\hat{\theta}\right)} \tag{13}$$

This test is compared to the chi square in case where the data distribution is not normal. In case where the data is normally distributed, the Wald test is given by:

$$W_N = \frac{\hat{\theta} - \theta_0}{se(\hat{\theta})} \tag{14}$$

where *se* is the standard error of the MLE estimate which is given by:

$$se = \frac{1}{\sqrt{I_n(MLE)}} \tag{15}$$

where I_n is the Fisher information (Harell, 2001, Fears *et. al.*, 1996, Engle, 1983, and Agresti, 2002). The finding of the Wald test in Table 4 shows that there is no significant difference between the arithmetic mean and the estimated mean. The practical implication for stock price analysis is that the new LLE can provide a more accurate estimation.

Table 4. Result of the Wald Test under Conventional MLE and New LLE

Market	θ_0	θ_1	$W = \frac{(\hat{\theta} - \theta_0)^2}{\text{var}(\hat{\theta})}$	$\chi^2(60) = 79.10$
Dow	17,614.72	17,708.73	0.00	Not significant
SP500	2,080.28	2,095.27	0.00	Not significant
NASDAQ	5,036.54	5,037.91	0.00	Not significant
DAX	4,927.64	4,896.49	0.00	Not significant
CAC40	11,117.15	11,103.28	0.00	Not significant

4.4 Lagrange Multiplier (Score Statistic)

The Langrange multiplier test is also called the score test. The score test had been explained by several authors, such as Bera (2001), Lehman and Casella (1998), Engle (1983), and Cook and Demets (2007). The score test is more appropriate where the deviation between $\hat{\theta}$ and θ is small; this is the case of the adjusted log likelihood proposed by this paper. The score test is given by:

$$U(\theta) = \frac{\partial \log L(\theta \mid X)}{\partial \theta} \tag{16}$$

The null hypothesis is $\theta = \theta_0$. If the null hypothesis cannot be rejected, the data is treated as chi square distribution. The test statistic is given by:

$$S(\theta_0) = \frac{U(\theta_0)^2}{I(\theta_0)} \tag{17}$$

where $I(\theta_0)$ is the Fisher information or $I(\theta_0) = -E\left[\frac{\partial^2}{\partial \theta^2} \log L(X \mid \theta) \mid \theta\right]$. For normally distributed data, the score test is given by:

$$S^*(\theta) = \sqrt{S(\theta)} \tag{18}$$

The result of the score statistic shows that the null hypothesis $\theta = \theta_0$ is true for all markets under 95% confidence interval. Only Shanghai index shows a significant difference between the actual and estimated value. Under this test, the new LLE could show 10 our of 10 cases in accuracy.

Table 5. Result of the Score Statistic Test under Conventional MLE and New LLE

Market	θ_0	θ_1	$\theta = \theta_0$	$\chi^2(60) = 79.10$
Dow	17,614.72	17,708.73	-0.5%	Not significant
SP500	2,080.28	2,095.27	-0.7%	Not significant
NASDAQ	5,036.54	5,037.91	0.0%	Not significant
DAX	4,927.64	4,896.49	0.6%	Not significant
CAC40	11,117.15	11,103.28	0.1%	Not significant

4.5 A Proposed Test for New LLE under Chi Square

This paper proposed a test statistic for the new LLE method by using chi square statistic in form:

$$\chi^2 = \left. \frac{(n-1)S^2_{\ln L(X)}}{\sigma^2_{E[X]}} \right|_{df=n-1} \qquad (19)$$

where n is the sample size; $S^2_{\ln L(X)}$ is the variance of the new $\ln L(X)$ under equation (18), and $\sigma^2_{E[X]}$ is the estimated variance of the $E[X]$ in (19). The critical value for the null hypothesis is read from the chi square table at degree of freedom $n-1$. Under this proposed test statistic, we are able to achieve more consistent result than the score statistic in Table 5.

Table 6. Result of the LLE Ratio Test under New LLE

Market	$S^2_{\ln L(X)}$	$\sigma^2_{E[X]}$	$\frac{(n-1)S^2_{\ln L(X)}}{\sigma^2_{E[X]}}$	$\chi^2(60) = 79.10$
Dow	308,045.00	1,834,128,509.00	0.01	Not significant
SP500	4,505.00	20,287,376.00	0.01	Not significant
NASDAQ	11,268.00	94,746,147.00	0.01	Not significant
DAX	51,553.00	606,504,707.00	0.01	Not significant
CAC40	15,654.00	143,094,150.00	0.01	Not significant

4.6 Out-of-Sample Test

Findings made from in-sample testing may not be reproduced when an out-of-sample test is conducted (Inuoe & Lutz, 2002). An out-of-sample test is the re-testing of the claimed made by a prior empirical whose conclusion was reach by using in-sample testing. In some cases, the out-of-sample data comes from the original sample where it has been split (Hansen & Timmermann, 2012). In the present case, the out-of-sample test employs a set of data different from the sample use for the hypothesis testing. The out-of-sample test consists of the daily indices of 10 major stock markets: (i) NASDAQ; (ii) DOW; (iii) SP500; and (iv) CAC40. The data was taken from a period of 60 days between June and August 2015.

Table 7. Out-of-Sample Test for Estimated Value under Old and New LLE Methods

Market	Max X_1	Min X_2	* X_3	Mean $\bar{X}$	S	μ
Dow	18,144.07	17,515.42	17,829.75	17,829.75	314.32	17,299.84
SP500	2,124.20	2,046.68	2,085.44	2,085.44	38.76	2,020.10
NASDAQ	5,160.09	4,909.76	5,034.93	5,034.93	125.16	4,823.91
DAX	11,542.54	10,676.78	11,109.66	11,109.66	432.88	10,379.88
CAC40	5,059.17	4,604.64	4,831.91	4,831.91	227.26	4,448.77

*((X1+X2)/2) = X_3.

The result of the Monte Carlo for the out-of-sample test follows the following decision rule: $H(0): \Pr\left[\frac{\bar{X}-\mu}{S/\sqrt{n}}\right] \le \Phi\left(\frac{3\sigma}{\sqrt{n}}\right)$, otherwise H(A) where S is the pool sample standard deviation. The result of the test is summarized in Table 7 below.

Table 8. Monte Carlo Simulation for 5 Major Stock Markets in Out-of-Sample Test

Market	$\bar{X}$	μ	$H(A): \Phi\left(\frac{\bar{X}-\mu}{S/\sqrt{n}}\right)$	$H(0): \Phi\left(\frac{3\sigma}{\sqrt{n}}\right)$
DOW	17,829.75	17,299.84	0.841	0.998
SP500	2,085.44	2,020.10	0.841	0.998
NASDAQ	5,034.93	4,823.91	0.841	0.998
DAX	11,109.66	10,379.88	0.841	0.998
CAC40	4,831.91	4,448.77	0.841	0.998

*Monte Carlo iteration: *n = 217*.

The result of the test shows that the null hypothesis cannot be rejected. The Monte Carlo simulation of 271 iterations produces estimated value within the range of confidence interval at 99.8%. In terms of stock price analysis, this out-of-sample test confirms that the proposed estimating method is generalizable, i.e. apply n situation outside of the empirical data. Recall that the empirical data used in this paper consists of stock prices of 100 companies in the SET100 index. The out-of-sample tests uses data from a different source: indices from 5 major stock markets. The practical implication of this confirmation is that the new log likelihood estimator is generalizable.

5. CONCLUSION

The New Log-Likelihood Estimation (NLLE) function presented in this paper is a novel discovery. It may serve as a better tool for risk management and value estimator. This innovation is a contribution to the field because they fill the gap in the literature and have practical utility. Beyond stock price analysis, NLLE also has general applications in other fields in natural and social sciences.

REFERENCES

DATA: The SET100 data used for this research may be obtained at the share drive at the following page link:
https://drive.google.com/file/d/0B772RxSNNsoVaURhMjMtOEFzN2M/view?usp=sharing

Agresti, Alan (2002). Categorical Data Analysis, (2nd ed.). Wiley. p. 232. ISBN 0471360937.

Bera, Anil K.; Bilias, Yannis (2001). "Rao's score, Neyman's C(α) and Silvey's LM tests: An essay on historical developments and some new results." Journal of Statistical Planning and Inference **97**: 9–44. doi:10.1016/S0378-3758(00)00343-8.

Casella, George; Berger, Roger L. (2001). Statistical Inference (Second ed.). ISBN 0-534-24312-6; p. 375.

Cook, T. D.; DeMets, D. L., eds. (2007). Introduction to Statistical Methods for Clinical Trials. Chapman and Hall. pp. 296–297. ISBN 1-58488-027-9.

Cramér, Harald (1946). Mathematical Methods of Statistics. Princeton, NJ: Princeton Univ. Press. ISBN 0-691-08004-6. OCLC 185436716.

Efron, Bradley (1981). "Maximum Likelihood and Decision Theory." *Technical Report No. 70* (Sep 1981). Stanford University: p. 1.

Engle, Robert F. (1983). "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics." In Intriligator, M. D.; and Griliches, Z. Handbook of Econometrics **II**. Elsevier. pp. 796–801. ISBN 978-0-444-86185-6.

Engle, Robert F (1984) . "Wald, Likelihood Ratio and Lagrange Multiplier tests in Econometrics," in *Handbook of Econometrics*, Volume II, Edited by Z. Griliches and M.D. Intriligator. Elsevier Science Publishers BV.

Everitt, Brian S. (2002). The Cambridge Dictionary of Statistics. Cambridge University Press. ISBN 0-521-81099-X.

Fears, Thomas R.; Benichou, Jacques; Gail, Mitchell H. (1996). "A reminder of the fallibility of the Wald statistic." The American Statistician **50** (3): 226–227. doi:10.1080/00031305.1996.10474384.

Fisher, R (1921). "On the Mathematical Foundations of Theoretical Statistics." Philosophical Transactions of the Royal Society of London. Series A **222**: 309–368.

Fisher, R.A., Tippett, L.H.C., (1928). "Limiting forms of the frequency distribution of the largest and smallest member of a sample." *Proc. Cambridge Philosophical Society* **24**:180–190.

Hansen, Peter Reinhard and Timmermann, Allan (2012). "Choice of Sample Split in Out-of-Sample Forecast Evaluation."
http://rady.ucsd.edu/docs/faculty/timmerman/samplesplitmining2012_feb07.pdf
Accessed; January 13, 2016.

Harrell, Frank E., Jr. (2001). "Sections 9.2, 9.3.1., 9.3.3 & 10.5" Regression modeling strategies. New York: Springer-Verlag. ISBN 0387952322.

Huelsenbeck, J. P.; Crandall, K. A. (1997). "Phylogeny Estimation and Hypothesis Testing Using Maximum Likelihood." Annual Review of Ecology and Systematics 28: 437–466. doi:10.1146/annurev.ecolsys.28.1.437.

Inoue, Atsushi and Kilian, Lutz, In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use? (December 2002). CEPR Discussion Paper No. 3671. Available at SSRN: https://ssrn.com/abstract=369560

Kroese, D. P.; Brereton, T.; Taimre, T.; Botev, Z. I. (2014). "Why the Monte Carlo method is so important today." *WIREs Comput. Stat.* **6**: 386–392. doi:10.1002/wics.1314.

Laplace, Pierre-Simon (1814). *Essai philosophique sur les probabilités*. Paris: Courcier. file:///C:/Documents%20and%20Settings/Administrator/Desktop/NEW%20BOOK/Essai%20philosophique%20sur%20les%20probabilités%20-%20Pierre%20Simon%20marquis%20de%20Laplace%20-%20Google%20Books.html

Lehmann, Erich L.; Casella, George (1998). Theory of Point Estimation, 2nd ed. Springer. ISBN 0-387-98502-6; p. 115.

Myung, In Jae (2003). "Tutorial on Maximum Likelihood Estimation." *Journal of mathematical Psychology*, **47**: 90-100.
Online: http://times.cs.uiuc.edu/course/410/note/mle.pdf

Neyman, Jerzy; Pearson, Egon S. (1933). "On the Problem of the Most Efficient Tests of Statistical Hypotheses." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **23** (694–706): 289–337.
Bibcode:1933RSPTA.231.289N. doi:10.1098/rsta.1933.0009. JSTOR 91247.

Nikulin, M.S. (2001). "Efficiency of a statistical procedure," in Hazewinkel, Michiel, Encyclopedia of Mathematics, Springer, ISBN 978-1-55608-010-4.

Data Classification and Distribution

Chanoknath Sutanapong★ & Louangrath, P.I. ★★

About the author

★Chanoknath Sutanapong is an independent researcher. She may be reached by email at: Chanoknath.sutanapong@gmail.com

★★ Louangrath, P.I. is an Assistant Professor in Business Administration at Bangkok University, Bangkok, Thailand. He could be reached by email at: Lecturepedia@gmail.com

ABSTRACT

The objective of this paper is to explain the four main types of data. The classification of data by type is important for statistical analysis. In particular, data classification is useful for quantitative research in social science. Data are defined as a quantitative measurement of qualitative fact. Data are classified into three types: quantitative, ordinal and nominal. Quantitative data are those that may be subject to mathematical operations: addition, subtraction, multiplication and division. Ordinal data are those that rank the values in a data set in an ascending order (from low to high) or from descending order (from high to low). Nominal data are those numbers or designation of value that is used for the purpose of identification. Nominal data cannot be subjected to mathematical operations. In addition to types, the data may also be classified according to their probability nature: (i) discrete data for discrete probability and (ii) continuous data for continuous probability.

Keywords: data types, quantitative data, nominal data, ordinal data

CITATION:

Sutanapong, C. and Louangrath, P.I. (2015). "Data Classification and Distribution" *Inter. J. Res. Methodol. Soc. Sci.*, Vol., 1, No. 2: pp. 36-67. (Apr. – Jun. 2015).

1.0 INTRODUCTION

1.1 Data classification

Data can be classified into two main types: (i) qualitative and (ii) quantitative. Qualitative data are those that are used for identification. No mathematical operations, such as multiplication and division may be allowed. Quantitative data are those that have non-arbitrary zero point and the data may be subjected to mathematical operations, such as addition, subtraction, multiplication and division. In statistical analysis, these two types of data are further divided into subcategories. For qualitative data, there are two subcategories: (1) nominal data, and (2) ordinal data. For quantitative data, there are also two further categories: (1) interval scale, and (2) ratio scale.

1.2 Nominal data and its central tendency

Nominal data is qualitative data; it is used to differentiate between items or subjects based on names or meta-categories, such as gender, nationality, ethnicity and language. There are three measures for central tendency: mean, median, and mode.

The *mean* is a common measure used for the measurement of the central tendency. The mean is the expected value of the set. A mathematical mean is the sum of all items divided by the number of items. In a given set of $X_i : (x_1, x_2, x_3, x_4, x_5)$, the means is $\bar{X} = (x_1 + x_2 + x_3 + x_4 + x_5) \div 5$. It is a rough estimate of the central value. There are some values in the set which are located above and below the mean. The measure of that dispersion is called variance and, it is standardized in a standard unit of measurement as the standard deviation.

The median is simply the midpoint of the set such that the probability of the value falling above or below that point is equal.

The *mode* is used as the measurement for central tendency. The mode is defined as the value that appears most in frequency in a data set. In discrete probability, the mod is the X at which the probability mass function is maximized.

1.3 Ordinal data and its central tendency

Ordinal is a second type of qualitative data. This is a ranked order: 1^{st}, 2^{nd}, 3^{rd},… The data can be sorted on ascending order or descending order. The ranking may be dichotomous in form, such as (Yes | No), or non-dichotomous, such as: *completely agree, most agree, most disagree,* and *completely disagree*

The *median* is used for the measurement of central tendency. The median is the middle-ranked. The mean $\bar{x}$ is not allowed because the ranked order: 1^{st}, 2^{nd}, 3^{rd},… data may not be added or divided. Therefore, $\bar{x}$ is not the measure for central tendency. However, in addition to the median, the mode may also be used as a measure for central tendency. IQ test, for instance, is ordinal data; it is ranked data. There is no measurement to quantify intelligence (Mussen, 1973).

1.4 Interval data and its central tendency

Interval data is quantitative. Each item may be different; however, no ratios among items are allowed. This means that no division may be performed. For example, Celsius scale is an interval scale. However, a ratio of Celsius is not allowed. One cannot say that 20 degree Celsius is "twice" as hot as 10 degree Celsius because zero degree Celsius is an arbitrary number, i.e. 0 Celsius is equal to -273.15 kelvin. The interval variable is sometimes referred to as "scaled variable." A mathematical term for scaled variable is *affine line*. In affine space, there is no point of origin.

The central tendency of an interval data is measured by the mode, median and arithmetic mean. The measurement of the dispersion include range and standard deviation. In interval scaled data, multiplication and division are not allowed. Since division is not allowed, studentized range and coefficient of variation may not be calculated. The point of origin is arbitrary defined; thus, the central moment may be determined. Coefficient of variation may not be determined since the mean is a moment about the origin.

1.5 Ratio data and its central tendency

Ratio scaled data is quantitative. The measurement is the estimation of the ratio between a magnitude of a continuous and a unit magnitude of the same kind (Michell, 1997). Ratio scale has unique and meaningful zero. Mass, length, duration, plane angle and energy are measured by ratio scale. Quantitative data that is obtained through a measurement is this type of data. Since zero is not arbitrary value; therefore, ratios are allowed. Multiplication and division are allowable mathematical operations.

The central tendency is measured by the mode, median, arithmetic mean are the basic measurements. In addition, geometric mean and harmonic mean may also be used. Studentized range and the coefficient of variation are used to measure dispersion.

Central tendency is the measure of the central value of a probability distribution (Weisberg, 1992). It is the average or the center point of a distribution. Common measures of the central tendency include the arithmetic mean, media, and mode. These three measurements are referred to as central tendency (Dodge, 2003).

Based on the law of large number (LLN), as the number of observation gets larger, there is a tendency for the estimated value to gravitate towards the mean of the group. This LLN also is another tool to understand the concept of central tendency theory.

In statistics, central tendency refers to the probability distribution of continuous data whose divider at 50/50 separating the lower range and upper range in equal area to be the mean. This mathematical mean is taken to define the central limit of the data distribution for estimating the value of the distribution, i.e. the expected value of the observation.

2.0 Probability distribution

A distribution is the fraction of individual events in relations to the whole number of observation. Adding all the individual events, the sum of the distribution is 1.0, i.e. each event is a proportion to the whole where the whole (of whatever is being observed) is 100% or simply 1.0.

A probability distribution is the probability of a subset of the possible outcome in relations to the entire observation. There are two types of probability based on the nature of the data: (i) discrete data produces discrete probability, and (ii) continuous data produces continuous probability function. Graphically, discrete probability produces a picture of a histogram where each data point stands alone and not connected to any other data point. Graphically, a continuous data set produces a continuous line or curve. The distribution is cumulative.

2.1 Discrete probability distribution

A data that is categorical, such as (Yes | No) is called discrete data. It is discrete because the observer must select one over the other. By selecting one choice, i.e. Yes, the other choice (No) is precluded. Two things are required: (i) event of interest which is generally defined as "Yes" or "Success." This event of interest is assigned a score; that score is generally a number 1.0. The other event which is not an event of interest is "No"; it is assigned a score of zero (0); (ii) the second requirement is the total known observation. The total number of observation must be known; if this number is not known, probability cannot be calculated.

For discrete probability, some times called binomial distribution, the probability of a specified score may be determined if the number of past success and the number of total event are known. The binomial distribution is given by:

$$P(X) = \frac{n!}{(n-X)!X!} p^X q^{n-X} \qquad (1)$$

where $p = \frac{s+1}{n+2}$ and $q = 1 - p$.

The test statistic is given by:

$$Z_{bin} = \frac{\frac{X}{n} - p}{\sqrt{\frac{pq}{n}}} \quad (2)$$

The critical value depends on the level of confidence. Generally, the level of confidence is set at 0.95; therefore, the critical value for Z is 1.65.

2.2 Continuous probability distribution

Continuous probability distribution is produced by continuous data. Continuous data is the type of data in which all data points are connected. The measurement of the probability of this type of data is in a form of probability density function. The data is cumulative and continuous. Graphically, it is represented as a continuous line or curve. A common form of representation is the Gaussian function which represents normal distribution:

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right] \quad (3)$$

Where …

x = actual sample event or observation;
μ = estimated mean of the assumed ideal population; and
σ = estimated population standard deviation.

The variables μ and σ are considered *inferential statistics* which may be determined through the t-equation and Z-equation:

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}} \quad \text{, solve for population mean } \mu = t\left(\frac{S}{\sqrt{n}}\right) - \bar{x}$$

and …

$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \quad \text{, solve for population standard deviation: } \sigma = \left(\frac{\bar{x}-\mu}{Z}\right)\sqrt{n}\,.$$

2.2.1 Normal distribution

Normal distribution is a function that explains the continuous probability of any given data point in the data set would fall between “two numbers.” These two numbers are the upper bound and lower bound with the mean as the reference point, i.e. $\bar{x} \pm S$. The function that produces a normal curve is called the Gaussian function:

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right] \quad (4)$$

The Gaussian function depends on three factors: the observation (x), the mean (μ) and the standard deviation (σ).

Probability density function (PDF) is defined as the density of a continuous variable. It is a function that gives the relative likelihood that a particular vale would take for a univariate, i.e. X. The function for the density is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{5}$$

where X is given. If more and more X are given then we will start to see a curve produced by f. The expected value for X is given by:

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx \tag{6}$$

Note that only continuous data has probability density function. Discrete data does not have probability density function. The requisite property of a probability density function is the cumulative distribution function (CDF). If there is no CDF, there cannot be PDF because this is a property that differentiates continuous random variable from discrete random variable.

Cumulative density function (CDF) is the area under the probability density function (PDF) from negative infinity up to point X. Recall that point X is a given value where we ask "what is the probability of X occurring? CDF exists only when the data variable is continuous and non-discrete. It cannot exists where the variable is discrete because in discrete data distribution, each data point is independent and does not connect to any other data points. Since the cumulative density is comprised of all area or region under the curve, cumulative density can only exists in continuous data form. The cumulative probability distribution function is given by:

$$F_X(x) = \int_{-\infty}^{x} f_X f(t)dt \tag{7}$$

where X = the entire data set of continuous data set $X_i : (-\infty, ..., \infty)$; x = a given data point that we want to determine its cumulative probability density up to that point; t = generally equated to time since each time we ask about the CDF of x we do so at a distinct time. We can answer about the cumulative probability of x one data point at a time since CDF is the cumulative of all probability up to that point x.

2.2.2 Standard normal distribution

Standard normal distribution is the Gaussian function for the normal distribution that has a mean of zero ($\mu = 0$) and standard deviation of one ($\sigma = 1$). This is a perfect bell shape curve. When we speak of normal distribution, we generally refer to the ideal form or standard normal distribution where the mean is equal to zero and the variance is one. The probability density function (PDF) of the standard normal distribution is given by:

$$\phi(x) = \frac{\exp\left(-\frac{1}{2}x^2\right)}{\sqrt{2\pi}} \tag{8}$$

Note that the factor $1/\sqrt{2\pi}$ is used to ensure that the entire area under the curve is equal to one, i.e. a unit distribution. The value of ½ in the exponent ensures a unit value or 1.0 variance and 1.00 standard deviation. The PDF is symmetric about zero; it means that it is a perfect mirror image of itself at zero mean ($\mu = 0$). The curve shifts direction or shows inflection points at -1 and +1.

2.2.3 General normal distribution

Every normal distribution is a version of standard normal distribution, meaning that it shares properties with the standard normal distribution. The mean (μ) is the reference point and the deviation from the mean is measured in units of standard deviation (σ). The probability density of the general normal distribution is given by:

$$f(x,\mu,\sigma)=\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) \qquad (9)$$

The factor of $1/\sigma$ is used as a scale to ensure that the probability density is equal to 1.0. If X is a general normal distribution, then:

$$Z=\frac{(x-\mu)}{\sigma} \qquad (10)$$

From the above statement, it also means that if Z is a standard normal distribution, then:

$$X = Z\sigma + \mu \qquad (11)$$

where the distribution of Z has the expected value equal to μ and standard deviation σ. Every normal distribution is an exponent of a *quadratic* function:

$$f(x)=e^{ax^2+bx+c} \qquad (12)$$

The property of the exponent may be summarized as follows:

$a < 0$	The a is negative.
$c = -\ln(-4\pi)/2$	The constant term
$\mu = -\frac{b}{a}$	The mean is a negative ratio of b/a
$\sigma^2 = -\frac{1}{2}a$	The variance is negative ½ of a.

For standard normal distribution, the properties are:

$a = -\frac{1}{2}$,

$b = 0$

$c = -\ln(2\pi)/2$

The *T-distribution* is used to analyze normal distribution of a sample. The sample is assumed to have normal distribution. Normal distribution is denoted as $N(0,1)$, i.e. identical, independent, distribution (IID) with mean of zero and variance of 1.0. Recall that a sample is a portion of a population taken with sample size of n from population where in the size of the population may be known or unknown or non-finite. This is the first scenario of normal distribution. The critical value of the t-distribution is given by the t-equation:

$t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$ (the value of t is given in the T-Table).

The *Z-distribution* is a normal distribution of a population. When the population (not the sample) is the unit of analysis, use the Z-table. All properties and assumptions of normal distribution used in the t-distribution scenario are applicable. The population distribution is explained by the standard score equation for the normal distribution of the population. The Z-equation is given by:

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$ (the value of Z is given in the Z-Table).

3.0 NON-NORMAL DISTRIBUTIONS

One approach of differentiate data characteristic is to verify whether the data is normally distributed. A normal curve manifests at least five distinct characteristics. First, *Symmetry* around the mean where the mean, median and mode are equal. Second, the distribution curve is *unimodal*. Third, the area under the curve is *unity*. Fourth, there is an *inflection* point at +/- 1 standard unit about the mean. Lastly, the density of the curve is *log-concave*. If the distribution curve of the data breaks these characteristics, it is considered not normally distributed. One short hand indicator for testing the data's normality are skewness and kurtosis. A normal distribution, skewness is 0 and kurtosis is less than 3. Skewness and kurtosis are calculated by:

$$SKEW = \frac{n}{(n-1)(n-2)} \sum \left(\frac{X_i - \bar{X}}{S} \right)^3 \quad (13)$$

$$KURT = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{X_i - \bar{X}}{S} \right)^4 \quad (14)$$

If the kurtosis exceeds 3.0, the data does not have normal distribution. The expected value for skewness is zero; however, there is a problem to conclude when skewness and kurtosis leads to a different conclusion. For stance, kurtosis is less than 3, but skewness is non-zero. In such a case, we have a conflicting conclusion. The solution to this apparent conflict is to convert skewness to a Z score and reads the percentage probability whether it exceeds a certain threshold, i.e. 5% error. This reconciliation may be accomplished by D'Agostino's K square statistic (D'Agostino, 1970; D'agostino *et al.*, 1990).

The K square statistic or Omnibus test is used to test whether the non-zero skew conforms to normal distribution (D'Agostino, Belanger, and D'Agostino, 1990). With known skewness and kurtosis, determine $Z_1(g_1)$ and $Z_2(g_2)$, then calculate K squared. If K sq. ≤ 2.0, it means the data is normally distributed. If K sq. > 2.0, it means the data is not normally distributed.

$$K^2 = Z_1(g_1)^2 + Z_2(g_2)^2 \quad (15)$$

where $Z_1(g_1) = \delta a \sinh\left(\frac{g_1}{\alpha \sqrt{\mu_2}} \right)$

$$W^2 = \left(\sqrt{2\gamma_2 + 4}\right) - 1$$

$$\delta = \frac{1}{\sqrt{\ln W}}$$

$$\alpha^2 = \frac{2}{W^2 - 1}$$

$$\mu_{1(g1)} = 0$$

$$\mu_{2(g1)} = \frac{6(n-2)}{(n+1)(n+3)}$$

$$Z_2(g_2) = \sqrt{\frac{9A}{2}}\left\{1 - \frac{2}{9A} - \left(\frac{1 - \frac{2}{A}}{1 + \left(\frac{g_2 - \mu_1}{\sqrt{\mu_2}}\right)\sqrt{\frac{2}{(A-4)}}}\right)^{1/3}\right\}$$

$$A = 6 + \frac{8}{\gamma_1}\left(\frac{2}{\gamma_1} + \sqrt{1 + \frac{4}{\gamma_1^2}}\right)$$

$$\mu_{1(g2)} = -\frac{6}{n+1}$$

$$\mu_{2(g2)} = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}$$

Table 1. K square statistic

Sample size	Expected value	Standard deviation	95% quantile
20	1.971	2.339	6.373
50	2.017	2.308	6.339
100	2.026	2.267	6.271
250	2.009	2.174	6.129
500	2.012	2.113	6.063
1000	2.009	2.062	6.038
$\chi^2(df = 2)$	2.000	2.000	5.991

3.1 Chi square distribution

When the sample size is small, the distribution of the data will not be normally distributed. In order to determine the goodness-of-fit, we need to compare our small sample to an ideal normal distribution. What does that mean? It means that with the small sample that we have, the data is not normally distributed. However, assume that the data is large enough then it would have been normally distributed. This "normal distribution," through assumption, is then compared to the assumed unit normal distribution (standard distribution) and compare our presumed "normal had we had large enough data" to the standard normal distribution and see how does our distribution fit to the standard one. If it is closely fit, we say it would meet the requirement of goodness-of-fit.

The goodness-of-fit under chi-square test is given by:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \tag{16}$$

Recall that s^2 is the sample variance which is given by:

$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$ and σ^2 population variance comes from $\sigma = \sqrt{\sigma^2}$ where:
$\sigma = \left(\frac{\overline{x} - \mu}{Z}\right)\sqrt{n}$ from the Z-equation: $Z = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}}$. The chi-square test statistic for goodness-of-fit is read from a chi-square table.

The chi-square equation is also given as:

$$\chi^2 = \sum_{i=1}^{k}\frac{\left(O_i - E_i\right)^2}{E_i} \qquad (17)$$

where ..

O_i = observed frequency;

E_i = expected frequency.

The critical value for the chi-square shown in the table depends on the reading of the degrees of freedom. The degree of freedom is determined by: $df = (K - 1 - 2)$ where K is the number of classes. If the theoretical distribution contains m parameters as in $K : \{1, 2, ..., m\}$, then the degree of freedom may be defined as $df = (K - 1 - m)$. This degree of freedom is used to read off the critical value in the chi-square table. The degree of freedom is located in the first column of the chi-square table.

If the critical value for the chi-square is less than the critical value given by the chi-square table, it is considered good. It is said that the data or model provides a good fit. “Good fit” of what? It is a goodness-of-fit that fits into the mantel of the standard normal distribution. The purpose of the chi-square is to use the small sample to fit the normal distribution “had the sample been the size that would have been adequate to produce a normal distribution.” If the observed critical value is larger than the expected value then it is said that the data or model does not fit the normal distribution. It can be said that the data is “truly non-normal” or significantly different from the normal distribution.

3.2 Poisson distribution

Recall that when the sample size is small and produces a non-normal distribution, the chi-square distribution is used to compare the assume distribution had the sample been increase to the magnitude that “would have produced a normal distribution.” The chi-square situation involves one sample. In case where there are two small samples and both of which are non-normally distribute and are taken at different time intervals, chi-square study would have be able to accommodate our analysis. The Poisson distribution analysis accommodates this second scenario: two chi-square distribution with time interval.

The Poisson distribution is a distribution of a discrete data. A discrete random variable X is said to have a Poisson distribution with parameter lambda: λ where $\lambda > 0$ if $k = 1, 2, 3, ...$ The probability mass function (PMF) is given by:

$$f(k; \lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad (18)$$

where $e = 2.71828$ and $k!$ is the factorial of k. The value of lambda is $\lambda = E(X) = Var(X)$.

Although Poisson distribution is used with discrete random variable, the Poisson itself can be counted---that is, two Poisson distribution can be counted and compared. Recall that the simple Poisson distribution involves two binomial distribution comparing counts from two different time frames.

The count of each Poisson is represented as N_1 and N_2 taken from two time frames t_1 and t_2. The average of the two frequencies are:

$$R_1 = \frac{N_1}{t_1} \quad \text{and} \quad R_2 = \frac{N_2}{t_2} \tag{19}$$

It is assumed that the assumed frequencies are equal. The test statistic for the equality is given by:

$$Z = \frac{R_1 - R_2}{\sqrt{\frac{R_1}{t_1} + \frac{R_2}{t_2}}} \tag{20}$$

The hypothesis statement follows: $H_0 : R_1 = R_2$. The null hypothesis argues that both frequencies have the same average, i.e. there is no significant difference. $H_A : R_1 \neq R_2$. The alternative hypothesis argues that the two average frequencies are not the same, i.e. they are statistically significantly different.

Generally, this type of test is used to prove changes after a stimulus is introduced into the system, i.e. new procedures or new policy, and a measurement is taken to prove whether the stimulus causes any changes or make the system respond by comparing the data from two time period. This type of comparison study is useful in empirical research. The phrase "comparison study" is not a grammatical error. The phrase is used when two data sets are compared. It is incorrect to use the term "comparative study." Comparison focuses on a certain characteristic of property of the data. Comparative study involves the entire set or all characteristics of the set.

3.3 F Distribution

Recall that a population study is tested by the Z-equation and the critical value for the distribution is given by the Z-table. Where the study involves two populations, the F-table is used. The two populations or groups may have different sizes. Different sizes imply different degrees of freedom. Recall that the degree of freedom is defined as $df = n - 1$. Thus, the F-table is read by using the degrees of freedom from each group.

F-distribution is a continuous probability distribution. It is used in the analysis of variance. In the t-distribution and Z-distribution, the means are used; therefore, those two tests are called means analysis. The variance is the shape of the distribution curve. Therefore, F-distribution is the comparison study of the two curves via their variances or the shape of the curves. The test statistic used to verify whether the two populations are significantly different is given by:

$$F = \frac{\left(S_1^2 / \sigma_1^2\right)}{\left(S_2^2 / \sigma_2^2\right)} \tag{21}$$

where $S_1^2 = \frac{\sum (x_i - \overline{x})^2}{n_1 - 1}$, and $S_2^2 = \frac{\sum (y_i - \overline{y})^2}{n_2 - 1}$;

Assuming that there are two groups of data call $X_i : (x_1, x_2, ..., x_n)$ and $Y_i : (y_1, y_2, ..., y_n)$, the mean for each group is given by:

$$\overline{x} = \sum_{i=1}^{n} x_i \qquad \text{and } \overline{y} = \sum_{i=1}^{n} y_i .$$

If the distribution is normal, the variances $\sigma_1^2 = 1$ and $\sigma_2^2 = 1$. Therefore, F becomes:

$$F = \frac{S_1^2}{S_2^2} \qquad (22)$$

The degree of freedom is defined as $df(n_1 - 1, n_2 - 1)$ where $n_1 - 1$ is the numerator degree of freedom and $n_2 - 1$ is the denominator degree of freedom. The hypothesis statement follows: $H_0 : S_1^2 = S_2^2$. The null hypothesis argues that both variances are equal, i.e there is no significant difference. The alternative hypothesis is $H_A : S_1^2 \neq S_2^2$. The alternative hypothesis argues that the two variances are not the same, i.e. they are statistically significantly different.

In the F-table, the top row (from left to right marked as d1) is the numerator degree of freedom. The first column (top to bottom marked as d2) is the denominator degree of freedom. For example, if d1 = 6 and d2 = 5, the F-critical value is 4.95. *What does it mean?* Recall that the F-test measures the significance of the variance between two groups. If the critical value from the observed data is less than 4.95, it means that the difference between the two groups or populations is not significant. However, if the calculation for $F = S_1^2 / S_2^2$ is greater than 4.95, it means that the difference among these two populations is significant, i.e. they are real difference.

For a second example, d1 = 6 and d2 = 10, the critical value for F is 3.22. The null hypothesis argues that the difference in variance among the two populations is not statistically significantly different, i.e. $H_0 : F_{obs} < 3.22$ and the alternative argument (your argument) states that $H_A : F_{obs} > 3.22$. If you have the data of the two populations, calculate the variance comparison according to the formula $F = S_1^2 / S_2^2$ and reach a conclusion according to the decision rule of H_0 and H_A then reject or accept H_0 accordingly.

4.0 CONCLUSION

This paper is a foundational materials on statistics needed for quantitative research in social science. In this part 2, we explore three main types of data: quantitative, ordinal, and nominal, and their central tendency. In addition, we traced four common types of test statistics, namely Student T, Z, chi square and F tests. The Student T test is used for sample analysis. The Z test is used for population or expected value analysis. Both T and Z test requires the data distribution to be normal. The chi square test is a test of fitness. The fitness attempt is to fit the distribution of the empirical data to the assumed normal distribution. If the sample is not normally distributed, the chi square test is used. For two samples, which are not normally distributed, the F test is used. The F test verifies the fitness via ratio analysis of two samples in order to differentiate one from the other. If there is a signif9cant difference, it means that the two samples came from a different source or process. Conversely, if the difference between the two samples is not significant, it is concluded that both sample may have come from the same source or process.

REFERENCES

D'Agostino, Ralph B. (1970). "Transformation to normality of the null distribution of 1." *Biometrika*, **57** (3): 679–681. doi:10.1093/biomet/57.3.679. JSTOR 2334794.

D'Agostino, Ralph B.; Albert Belanger; Ralph B. D'Agostino, Jr (1990). "A suggestion for using powerful and informative tests of normality." *The American Statistician*. 44 (4): 316–321. https://web.archive.org/web/20120325140006/http://www.cee.mtu.edu/~vgriffis/CE%205620%20materials/CE5620%20Reading/DAgostino%20et%20al%20-%20normaility%20tests.pdf

Dodge, Y. (2003) The Oxford Dictionary of Statistical Terms, OUP for International Statistical Institute.ISBN 0-19-920613-9 (entry for "central tendency"); and Upton, G.; Cook, I. (2008) Oxford Dictionary of Statistics, OUP ISBN 978-0-19-954145-4 (entry for "central tendency").

Henk, Tijms (2004). Understanding Probability: Chance Rules in Everyday Life. Cambridge: Cambridge University Press. p. 169. ISBN 0-521-54036-4.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. British Journal of Psychology, 88, 355–383.

Mussen, Paul Henry (1973). Psychology: An Introduction. Lexington (MA): Heath. p. 363. ISBN 0-669-61382-7. "The I.Q. is essentially a rank; there are no true 'units' of intellectual ability."

Pólya, George (1920). "Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem" [On the central limit theorem of probability calculation and the problem of moments]. *Mathematische Zeitschrift* (in German). 8 (3–4): 171–181. doi:10.1007/BF01206525.
https://gdz.sub.uni-goettingen.de/id/PPN266833020_0008

Weisberg H.F. (1992) Central Tendency and Variability, Sage University Paper Series on Quantitative Applications in the Social Sciences, ISBN 0-8039-4007-6. p.2.

Economic Diversity of the ASEAN Countries

Louangrath, P.I. ★

About the author

★Louangrath, P.I. is an Assistant Professor in Business Administration at Bangkok University, Bangkok, Thailand. He could be reached by email at: Lecturepedia@gmail.com

ABSTRACT

This research examines economic identity of the ASEAN countries. It is proved that the ASEAN is a heterogeneous group lacking common economic identity. The method used in this research involves series of statistical tests. The data is comprised of economic data from the IMF's World Economic Outlook 2013. Four main economic issues are used: per capita GDP from 2004 to 2013, GDP growth rates from 2004 to 2013, HDI-2013 for ASEAN member countries, and Gini coefficient. Following the Dixon and Damgaard's approach, the Gini coefficient for the ASEAN region was calculated. The Anderson-Darling tests proved that the per capita GDP among the member countries are not normally distributed. This finding points to the lack of homogeneity within the group. Secondly, the paired means difference analysis of the per capita GDP and the GDP growth rates confirmed that there is a significant difference among the ASEAN countries. Thirdly, the HDI for 2013 shows significant difference among the ten countries ($t_d = 2.68$). While the Gini coefficient for the individual countries ranges from 35 to 47, the Gini coefficient for the ASEAN group is about 65. The Laplace trend test for 2004 to 2013 shows that there is a significant trend narrowing the gap of the regional Gini coefficient ($Z_{Laplace} = 3.38$). The Weibull regression also shows a negatively slope for ASEAN's Gini coefficient year-by-year trend. This research confirms that the ASEAN countries do not have a common economic identity. The empirical fact would make market integration under AEC-2015 a great challenge for stakeholders.

Key words: AEC-2015, Anderson-Darling test, ASEAN heterogeneity, economic identity, Gini coefficient, Human Development Index (HDI), Laplace trend test, and Weilbull distribution.

CITATION:

Louangrath, P.I. (2015). "Economic Diversity of the ASEAN Countries." *Inter. J. Res. Methodol. Soc. Sci.*, Vol., 1, No. 2: pp. 68-80. (Apr. – Jun. 2015).

1.0 INTRODUCTION

This paper asserts that there is an economic diversity among the ASEAN countries. Diversity is defined as differences in identity. We based the identity of the ASEAN countries on two economic indicators: GDP and HDI. Identity is defined as a sense of belonging. (Lee, 2000, p. 29). The shared sentiment among group members may be created on the basis on shared descent, language, sanctuaries, and customs (Leoussi and Grosby, 2006, p. p. 115). Common identity may be seen through ethnicity (Wallman, 1977, pp. 531-532). Some writers explained identity on the basis of nationalism (Rothi, 2005, pp. 135-155; and Miller, 1995). The introduction of the AEC-2015 introduces unique form of social identity. This new identity transcends national border to include 10 countries in the ASEAN region. The AEC wants to create a regional community on the basis of cultural, educational and economic integration. This new identity transcends the traditional notion of self-concept in relations to a social group (Turner, 1986, p. 237-253). It involves the acculturation of more than six hundred million people into adopting the idea of shared ideal of "one vision, one identity, one community."

Writers claimed that social identity is used as a means to manage intergroup behavior (Taifel and Turner, 1979, pp. 33-47, and Turner, 1999, p. 6-34). The goal of the AEC is economic integration; a creation of a common market. The creation of AEC's shared vision, identity and community is a means to facilitate regional economic integration. This paper examines whether there is a shared identity on the basis of economics in this new community. Four main factors are used as indicators for economic identity: (i) per capita GDP, (ii) GDP growth rate, (iii) HDI and (iv) intra-regional Gini coefficient. Lacking economic homogeneity, market integration would face an up hill battle. From the European Union's experience it was found that economic heterogeneity hinders the groups in realizing its full potential. (Hopner and Schafer, 2012, p. 19). In the proposed AEC, the blue print for integration exists. However, there are at least four countries: Cambodia, Laos, Myanmar and Vietnam, that are less developed in comparison to the other six original members. With economic heterogeneity, the AEC may repeat the EU's experience. Part of the EU's current financial crisis is the lack of economic homogeneity among group members.

The intended contribution of this research is to provide stakeholders with a new lens to look at the AEC. By verifying whether the AEC has an economic identity, stakeholders can better assess the probable outcome of policy implementation to achieve the goal of "one vision, one identity, one community."

2.0 DATA

The data used in this research comes from the World Economic Outlook 2013 published by the International Monetary Fund (IMF). Relevant economic data for the ASEAN countries from 2004 to 2013 were separated for analysis. These data include: (i) per capita GDP, and (ii) annual GDP growth rate. Additional data, such as Human development Index (HDI) came from the annual statistics produced by the United Nations Development Programme (UNDP) for the year 2013. The Gini coefficient for the region was calculated using Dixon and Damgaard approach.

Table 1. Annual GDP of 10 countries in ASEAN

Country	2010	2011	2012	2013	2014
Brunei	79,302.82	82,567.71	83,659.61	82,052.74	80,221.21
Cambodia	2,459.64	2,646.49	2,840.40	3,055.27	3,280.68
Indonesia	8,432.70	8,973.56	9,554.34	10,108.43	10,661.63
Laos	4,382.25	4,761.36	5,153.34	5,576.56	6,022.04
Malaysia	20,335.84	21,498.39	22,742.22	23,630.68	25,088.81
Myanmar	3,678.78	3,932.90	4,262.68	4,655.78	5,074.26
Philippines	5,550.36	5,773.74	6,122.30	6,545.99	6,953.29
Singapore	70,657.26	75,113.21	77,690.83	81,647.56	85,227.12
Thailand	13,187.95	13,513.55	14,690.31	15,252.06	15,596.54
Vietnam	4,395.52	4,716.98	5,000.76	5,300.32	5,657.25

Source: https://www.imf.org/external/pubs/ft/weo/2018/01/weodata/index.aspx

3.0 METHODOLOGY

Series of statistical tests were employed in order to identify and verify economic identity of the ASEAN countries as a group. Firstly, the Anderson-Darling test was used to test data distribution. The Anderson-Darling test is a statistic defined to improve the Kolmogorov-Smirnov test in the test of the distribution (Press, 1992, p. 621). It is a test used to confirm whether the data is normally distributed. Secondly, the paired means comparison study was used to verify the intra-group homogeneity among the ASEAN countries. This finding was verified by the paired t-test. The UNDP reports each country's HDI annually. This figured was used to determine whether the ten countries in the ASEAN are significantly different on the basis on HDI score. Group homogeneity on the basis of income distribution was also tested through Gini coefficient analysis. Finally, the Gini coefficient for the ASEAN was calculated over a period from 2004 to 2013. The Laplace trend test was use to verify the direction of the trend for the group's Gini coefficient. Throughout the tests, a confidence interval at 0.95 was used.

The economic data consists of 10 years from 2003 to 2014 inclusive. For purposes of minimum sample size, these ten years are considered adequate by using the Anderson-Darling test as the standard requiring $n > 5$. This 10 years period mirrors the period used for 10-years US treasury bond (O'Sullivan and Sheffrin, 2003, p. 197).

4.0 FINDINGS AND DISCUSSION

The distribution of the data for the ten years covering the period between 2004 and 2013 was verified by the Anderson–Darling test. The Anderson-darling test is a statistic defined to improve the Kolmogorov-Smirnov test in the test of the distribution (Press, 1992, p. 621). The test statistics for the AD test is given by:

$$AD = -n - S \quad (1)$$

where n is the sample size and S is given by:

$$S = \sum_{i=1}^{n} \frac{2i-1}{n}\left[\ln(F(Z)) + \ln(\ln(1 - F(Z)))\right] \quad (2)$$

The AD observed value may be adjusted by:

$$AD^* = AD\left(1 + \frac{0.752}{n} + \frac{2.25}{n^2}\right) \quad (3)$$

The theoretical value for the AD test at 95% confidence interval is 2.27

The result of the AD test is tabulated in Table 2. For the 10 years period, the GDP for the ASEAN countries are not normally distributed. We report the result of the AD test in two segments. First, the AD test result for the individual country's GDP between 2010 and 2014 (5 years period) is reported in Table 2. Second, the result of the AD test for the 10 countries as a group is reported in Table 3.

Table 2. Anderson-Darling test results for individual country GDP for FY2010-2014

Country	AD(obs)	AD*(95%)	Conclusion
Brunei	3.96	2.492*	Not normal
Cambodia	4.49	2.492	Not normal
Indonesia	4.37	2.492	Not normal
Laos	4.39	2.492	Not normal
Malaysia	4.40	2.492	Not normal
Myanmar	4.51	2.492	Not normal
Philippines	4.60	2.492	Not normal

Singapore	4.34	2.492	Not normal
Thailand	4.27	2.492	Not normal
Vietnam	4.37	2.492	Not normal

*See Appendix 1 for the theoretical value of the Anderson-Darling test.

The result of the Anderson-Darling test for the ASEAN as a single group for a period of 5 years shows that the distribution of the group's GDP is not normal. The annual AD test result for each year is tabulated in Table 2. The result of the group's non-normality of their GDP distribution leads us to speculate that the ASEAN countries are diverse. This diversity was later verified in Tables 3 and 4.

Table 2. Anderson-Darling test results for ASEAN 10 countries for FY2010-2014

ASEAN group	AD(obs)	AD*(95%)	Conclusion
FY2010	8.11	2.492	Not normal
FY2011	8.20	2.492	Not normal
FY2012	8.28	2.492	Not normal
FY2013	8.57	2.492	Not normal
FY2014	8.85	2.492	Not normal

The AD test is generally followed by the adjacency test in order to verify whether the data is a set of random events. The adjacency test is a tool to verify whether a data set is comprised of random numbers. (Bassein, 1996, pp. 482-490). Random number is defined as a number or string of numbers whose large set (i) creates or manifest an underlying distribution, and (ii) each number or element of the set is independent. Independent means that one number is not correlated to the successive number. The null hypothesis may be rejected if the test statistic lies outside of the lower and upper bound of the critical value. (Hart, 1942, pp. 445-7). The hypothesis statements are given by: $H_0 : L_{obs} < L_{0.05}$ and $H_A : L_{obs} > L_{0.05}$.

The critical value of L is given in a range of lower and upper value. If the test statistic value falls within this range, the null hypothesis cannot be rejected. Recall that the null hypothesis states that the pattern of the data is random. If the test statistic for the observation falls out of the range, i.e. lower than the lower bound or higher than the upper bound, the series is not random. The null hypothesis assumes random distribution. The purpose of the test is to reject this assumption. The null hypothesis is rejected if the observed value falls outside of the L-range. Adjacency test for randomness is accomplished by:

$$L_{n<25} = \frac{\sum_{i=1}^{n-1} \left(X_{i+1} - X_i \right)^2}{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2} \qquad (4)$$

There are ten countries in the ASEAN for five years period: $n = 5$. The null hypothesis value for the L-critical at 0.95 confidence interval is: $1.06 < L_{0.05} < 2.94$. The table below shows the result of the adjacency test for 10 countries in 5 years.

Table 3: Per Capita GDP of the ASEAN Countries are Random Events

Item	Year	L(obs)	L(0.05)	Result
1	FY2010	0.000009	1.06 < L < 2.94	Not Random

2	FY2011	0.000010	1.06 < L < 2.94	Not Random
3	FY2012	0.000010	1.06 < L < 2.94	Not Random
4	FY2013	0.000009	1.06 < L < 2.94	Not Random
5	FY2014	0.000009	1.06 < L < 2.94	Not Random

The null hypothesis is rejected. The per capita GDP among the ASEAN countries are not normally distributed and are non-random numbers. These two findings points to the possibility of intra-group heterogeneity; this lack of homogeneity may be evidence for deeper issues and challenges facing the AEC-2015 regionalization and the new social identifier of "one vision, one identity and one community" in the economic sphere may not easily be achieved.

4.1 Per capita GDP and the GDP Growth Rates among the ASEAN Countries

The purpose of analyzing the per capita GDP is to verify whether there is an intra-ASEAN homogeneity. This test is accomplished through two statistical tests. Firstly, the paired mean analysis is accomplished by d-bar analysis test (Rubin, 1973, pp. 159-183; Anderson, 1980, pp. 61-66; and Kupper *et al*., 1981, pp. 271-291). The d-bar analysis is given by:

$$t_d = \frac{\bar{d}}{S_d / \sqrt{n}} \tag{5}$$

Secondly, the conclusion reached in the d-bar analysis is verified by the paired t-test. (Goulden, 1956, pp. 50-55). These double-tests are used in order to avoid Type I error. The paired t-test is provided by:

$$T_{paired} = \left(\bar{X} - \bar{Y}\right)\sqrt{\frac{n(n-1)}{\sum\left(\hat{X}_i - \hat{Y}_i\right)}} \tag{6}$$

Alternatively, the paired means T test may be calculated by:

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{7}$$

The results of both tests are tabulated in Table 3 showing that there are significant difference among the ASEAN countries on the basis of per capita GDP and GDP growth rate. These findings show that there is a lack of homogeneity among the ASEAN countries.

Table 4: T value of paired Mean for Per Capita GDP of group and country

Country	FY2010	FY2011	FY2012	FY2013	FY2014
Brunei	5.39	5.39	5.21	5.16	5.11
Cambodia	(3.31)	(3.31)	(3.71)	(3.91)	(4.12)
Indonesia	(0.18)*	(0.18)*	(0.50)*	(0.65)*	(0.80)*
Laos	(1.77)	(1.77)	(2.14)	(2.31)	(2.49)
Malaysia	2.06	2.06	1.80	1.69	1.58*
Myanmar	(2.24)	(2.24)	(2.62)	(2.80)	(2.98)
Philippines	(1.30)*	(1.30)*	(1.66)	(1.82)	(1.99)
Singapore	5.27	5.27	5.08	5.04	4.99
Thailand	0.90*	0.90*	0.60 *	0.48 *	0.35*

Vietnam	(1.85)	(1.85)	(2.22)	(2.39)	(2.57)

*No significant difference from group mean.

The lack of homogeneity among the ASEAN countries may present greater challenges to the AEC-2015 integration. The idea of an ASEAN identity may become unrealistic. In order to minimize the intra-group difference on the basis of language, the ASEAN suggested the use of English as a common language. As for cultural and historical diversity within the group, ASEAN member countries celebrate its rich diversity. However, if the AEC should be a common market; ideally it should be economically homogeneous. Hemogeneity in regional economies ensures equality among members and inclusivity of growth as a single market. However, since the 10 countries in the ASEAN are economically diverse, inclusive growth may be a challenge that is not easily overcome.

4.2 Human Development Index Trend among the ASEAN Countries

The second indicator of economic identity for the ASEAN countries is HDI. Human Development Index measures the impact of economic policies on the quality of life (Davies and Quinlivan, 2006, pp. 868-876). HDI is not a year-to-year indicator. For that reason, 2013 is used in this paper because it represents the latest figure for HDI measurement (McGillivray and White, 2006, pp. 183-192). The HDI from 2010 to 2014 for the ASEAN countries is shown in Table 5.

Table 5. HDI for ASEAN countries FY2010 - FY2014

Country	FY2010	FY2011	FY2012	FY2013	FY2014
Brunei	0.846	0.852	0.860	0.863	0.864
Cambodia	0.533	0.540	0.546	0.553	0.558
Indonesia	0.662	0.669	0.677	0.682	0.686
Laos	0.542	0.554	0.563	0.573	0.582
Malaysia	0.774	0.776	0.779	0.783	0.787
Myanmar	0.526	0.533	0.540	0.547	0.552
Philippines	0.669	0.666	0.671	0.676	0.679
Singapore	0.911	0.917	0.920	0.922	0.924
Thailand	0.720	0.729	0.733	0.737	0.738
Vietnam	0.655	0.662	0.668	0.675	0.678

The ASEAN are divided into two groups. The old ASEAN members have siz countries: Brunei, Indonesia, Malaysia, Philippines, Singapore, and Thailand. The second group with the new members consists of Cambodia, Laos, Myanmar, and Vietnam. Using the paired T test, using equation7, we report the result of their significance of the difference in Table 6.

Table 6. Significance of difference in HDI for group 1 and group 2 in ASEAN

Country	FY2010	FY2011	FY2012	FY2013	FY2014
Mean1	0.76	0.77	0.77	0.78	0.78
μ1	0.68	0.68	0.69	0.70	0.70
Var1	0.01	0.01	0.01	0.01	0.01
Mean2	0.56	0.57	0.58	0.59	0.59
μ2	0.51	0.51	0.52	0.53	0.54
Var2	0.004	0.004	0.004	0.004	0.003
T(obs)	0.43	0.45	0.45	0.45	0.46
T*(95%)	1.64	1.64	1.64	1.64	1.64
pValue	0.238	0.22	0.22	0.22	0.22
Conclude	Not sig.	Not sig.	Not sig.	Not sig.	Not sig.

*The procedure on how to calculate the pValue is given in Appendix 2.

The result of the calculation for the significant difference of HDI among countries in Table 6 shows that there is no significant difference in the 5 years period of the study. This finding is consistent with the findings of no significance in the GDp analysis.

As part of the new social identity intended by the AEC, social development is keyed. As a measurement of social development, HDI comparison helps stakeholder to see where member countries stand with respect to other group members. The above result shows that in terms of human development, the ASEAN countries have homogeneity. This finding further underscores the importance of common grounds for the 10 countries when it comes to economic identity for the group. It may be easy to promote market integration and a common market where member countries are on similar scale of development. The ideal of inclusive growth could be achieved if members of the group are uniform or similar in their development. This homogeneity in HDI does not call into question the vision of "one vision, one identity, one community."

4.3 Intra-ASEAN Gini Coefficient as Evidence of Economic Heterogeneity

The fourth factor used to indicate economic identity for the ASEAN is the Gini coefficient. The Gini coefficient is the measure of statistical dispersion of income distribution within a population. A Gini coefficient with a value of 1.00 means a maximum inequality within the group. It was first proposed by Gini as a measure for income inequality (Gini, 1936). As a measure of inequality, Gini coefficient is widely used in various fields of social science (Sadras and Bongiovanni, 2004, pp. 303-310). Two approaches to Gini coefficient calculation are currently in use: unordered data and ranked data. Gini coefficient for unordered data:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i - x_j\right|}{2n^2\mu} \qquad (8)$$

For increasing size (ascending order) data:

$$G = \frac{\sum_{i=1}^{n}\left(2i - n - 1\right)\hat{x}_i}{n^2\mu} \qquad (9)$$

This paper adopts equation (9) to determine the regional Gini coefficient. In so doing, the current per capita income for each country is used. It is assumed that each country is an element of a group or population. The number of observation for ASEAN is: $n = 10$. For the period from 2004 to 2013, the Gini coefficients for the ASEAN group are: 0.70, 0.69, 0.69, 0.69, 0.67, 0.67, 0.67, 0.67, 0.66, 0.65 and 0.65 while for the individual countries, the Gini coefficient provided by the UNDP ranges from 0.35 to 0.47 for the year 2014. This discrepancy between country and region points to a potential development problem for the AEC.

If the Gini coefficient is used as an indicator of economic equality, domestically each country fairs better than being part of the AEC. A higher Gini coefficient among the group also implied that there are rich and poor countries within the proposed economic community. With unbalanced economic development in unequal income distribution, the ideal of "one vision, one identity, one community" presents a greater challenge to achieve. As for the concept of shared identity, the economic divide among member countries would hinder uniformity. This finding should alert concerns for stakeholders and policy makers in the AEC.

The result of equation (9) shows the gulf in income distribution. The Gini coefficient from 2010 to 2014 provides an adequate data for a trend test verifying whether there is an improving or deteriorating trend. The Laplace trend test is used to verify whether there is an improving or deteriorating trend (Bohoris, 1996, pp. 18-28). The Laplace trend test is given by:

$$Z_{Laplace} = \sqrt{12r}\left(\sum_{i=1}^{r}\left(T_i - \frac{T_{end}}{2}\right)\right) \tag{10}$$

The result for the calculation under equation (10) is $Z_{obs} = 3.38$ where the theoretical value is $Z_{0.95} = 1.65$. There is a significant decreasing trend for the Gini coefficient using the period between 2004 and 2013 as the basis. This finding suggests that there is a positive prospect for the ASEAN countries because the income gap among the 10 countries significantly decreases.

4.4 Weibull Analysis for the AEC's System Failure Forecast

Although the finding from the Laplace trend test shows that there is a significant reduction trend narrowing the Gini coefficient gap among the ASEAN countries, further question must be asked: what is the probability that such a gap-closing trend will continue to occur? In order to answer this question, it is necessary to engage the Weibull test.

The Weibull test is a tool to predict failure. With a series of observation of $x_1, x_2, ..., x_n$ where each observation is observed in discrete time $t_1, t_2, ..., t_n$ such that the event series observed is in a form $((x_1 : t_1), (x_2 : t_1), ..., (x_n : t_1))$. This series of events is compared to the expected standard below which is considered as failure. Call that expected value $\hat{x}$. The time failure may be created by the following conditions: (i) if the sample size is less than 100, use the following formula:

$$F(t) = \frac{(i + 0.30)}{(n + 0.40)} \tag{11}$$

where i = the i[th] number of event in $((x_1 : t_1), (x_2 : t_1), ..., (x_n : t_1))$ and n = sample size.

As part of the Weibull distribution test, it is necessary to general the XY arrays for the Weibull distribution. Recall that the series $((x_1 : t_1), (x_2 : t_1), ..., (x_n : t_1))$ are observed events. To make sense of these observations in terms of failure rate, it is necessary to generate the Weibull plot to obtain the values for X and Y in order to run a regression for the Weibull distribution. The X-array may be generated by the following formula:

$$X_i = \ln\left(\ln\left(\frac{1}{1 - F(t)}\right)\right) \tag{12}$$

Thus from the observations series $((x_1 : t_1), (x_2 : t_1), ..., (x_n : t_1))$, a separate array of X_i is generated to correspond with each i in $((x_1 : t_1), (x_2 : t_1), ..., (x_n : t_1))$. With known X_i, the dependent Y_i array is generated by

$$Y_i = \ln(x_i) \tag{13}$$

With two arrays generated by (11) and (12), it is possible to run a Weibull regression from known Weibull plot:

$$\left.\begin{array}{l} x_1 : t_1 \to X_1 \\ x_2 : t_2 \to X_2 \\ \vdots \\ x_n : t_n \to X_n \end{array}\right\} \to X_i = \ln\left(\ln\left(\frac{1}{1-F(t)}\right)\right) \quad \text{and} \quad \left.\begin{array}{l} x_1 : t_1 \to Y_1 \\ x_2 : t_2 \to Y_2 \\ \vdots \\ x_n : t_n \to Y_n \end{array}\right\} \to \ln(x_i)$$

From the above arrays, a linear regression equation in a form of $Y_w = a + bX$ may be derived. In this case, the Weibull linear equation for the Gini coefficient trend is $Y_w = -0.41 + 0.021X$ with the coefficient of determination of $R^2 = 0.89$. Other relevant Weibull statistics include: $\beta = -47.34$, $\eta = 0.67$; and the Weibull reliability is $R_w = 0.37$. The Weibull reliability measures the reliability of the system; in this case, it is 37% reliable or the probability that the Gini coefficient gap will close is 37%. The finding under the Laplace's formula is also confirmed by Weibull's beta. In Weibull statistics, if the value of beta is less than zero, it means that the failure (Gini coefficient gap) is decreasing (gap closing) in the following periods: $t_i : (t_{n+1}, ..., T_N)$. The Weibull's eta value indicates the level of failure at the Y-intercept. In this case, eta is 0.67.

The finding under Weibull system failure analysis suggests that there is a decreasing trend in Gini coefficient; the gap in income distribution will narrow in relations with time. Out of the various factors examined as the indicator for economic identity for the AEC, the decreasing trend of the Gini coefficient appears to provide a common ground upon which all countries could agree. As a common market and an economic community, this finding is an important indicator that there is a commonality for which the 10 countries could look to as a shared identity in the new economy.

5.0 CONCLUSION

This research presents the concept of economic identity for the ASEAN countries. Four factors were used as the indicators for economic identities: per capita GDP, GDP growth rate, HDI and Gini Coefficient. Among these four factors, it was found there only the Gini coefficient trend show evidence of homogeneity among the group. On the basis of per capita GDP, GDP growth, and HDI, the ASEAN countries are economically heterogenous; however, as a group both grop1 and group2 are homogenous. On individual country basis, this heterogeneity would present greater challenge for the ASEAN members to achieve the AEC's shared ideal of "one vision, one identity, one community." However, as a group the ASEAN may integrate well. The intended contribution of this research is to provide practical assessment of the claim of shared identity among the ASEAN countries on the basis of economic indicators. The 10 countries are economically diverse. This economic diversity will hinder the prospect of smooth transition into a common market under the aegis of the AEC. In economic terms, the ASEAN lacks shared identity.

REFERENCES

Anderson, Dallas W.; Kish, Leslie; Cornell, Richard G. (1980). "On Stratification, Grouping and Matching." *Scandinavian Journal of Statistics* (Blackwell Publishing) **7** (2): 61–66.

Athena S. Leoussi, Steven Grosby (2006). *Nationalism and Ethnosymbolism: History, Culture and Ethnicity in the Formation of Nations*, Edinburgh University Press, 2006, p. 115.

Bassein, A. (1996). "A Sampler of Randomness." Amer. Math. Monthly **103**, 483-490.

Beyer, W. H. (1987). *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 491-492.

Bohoris, G.A. (1996), "A trend testing in reliability engineering." *International Journal of Quality & Reliability Management*, Vol. 13, No.6, pp.18-28.

Dagum, C. (1980). "The Generation and Distribution of Income, the Lorenz Curve and the Gini Ratio." *Écon. Appl.* **33**, 327-367.

Damgaard, C. and Weiner, J. (2000). "Describing Inequality in Plant Size or Fecundity." *Ecology* **81**, 1139-1142.

Davies, A.; Quinlivan, G. (2006). "A Panel Data Analysis of the Impact of Trade on Human Development." *Journal of Socio-Economics* **35** (5): 868–876.

Dixon, P. M.; Weiner, J.; Mitchell-Olds, T.; and Woodley, R. (1987). "Bootstrapping the Gini Coefficient of Inequality." *Ecology* **68**, 1548-1551.

Dixon, P. M.; Weiner, J.; Mitchell-Olds, T.; and Woodley, R. (1988). "Erratum to 'Bootstrapping the Gini Coefficient of Inequality.'" *Ecology* **69**, 1307.

Gini, C. (1912). "Variabilitá e mutabilita." Reprinted in *Memorie di metodologia statistica* (Ed. E. Pizetti and T. Salvemini.) Rome: Libreria Eredi Virgilio Veschi, 1955.

Hart, B.I. (1942). "Significance Level for the Mean Square Successive Difference to the Variance." *Annals of Mathematical Statistics*, **13**: 445-7.

Lorenz, M. O. (1905). "Methods for Measuring the Concentration of Wealth." *Amer. Stat. Assoc.* **9**, 209-219.

Glasser, G. J. (1962). "Variance Formulas for the Mean Difference and Coefficient of Concentration." *J. Amer. Stat. Assoc.* **57**, 648-654.

Goulden, C. H. (1956). *Methods of Statistical Analysis, 2nd ed.* New York: Wiley, pp. 50-55.

Höpner, Martin and Schäfer, Armin (2012). Integration among Unequals: How the Heterogeneity of European Varieties of Capitalism Shapes the Social and Democratic Potential of the EU." Discussion Paper No. 12/5. Max Planck Institute for the Study of Societies, Cologne, Germany. July 2012. ISSN 0944-2073 (print). ISSN 1864-4325 (online).

Kupper, Lawrence L.; Karon, John M.; Kleinbaum, David G.; Morgenstern, Hal; Lewis, Donald K. (1981). "Matching in Epidemiologic Studies: Validity and Efficiency Considerations." *Biometrics* (International Biometric Society) **37** (2): 271–291.

Lee, Yoonmi (2000). *Modern Education, Textbooks, and the Image of the Nation: Politics and Modernization and Nationalism in Korean Education: 1880-1910*. Routledge (published 2012). p. 29. ISBN 9781136600791.

McGillivray, Mark; White, Howard (2006). "Measuring development? The UNDP's human development index." *Journal of International Development* **5**(2): 183–192.

Miller, David (1995). *On Nationality*. Oxford: Oxford University Press. ISBN 0-19-828047-5.

O'Sullivan, Arthur; Sheffrin, Steven M. (2003). *Economics: Principles in action*. Upper Saddle River, New Jersey 07458: Pearson Prentice Hall. pp. 197, 507. ISBN 0-13-063085-3.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. (1992). *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 621.

Rothi, Despina et al. (2005). National attachment and patriotism in a European nation: A British study. Political Psychology, 26, 135 - 155.

Rubin, Donald B. (1973). "Matching to Remove Bias in Observational Studies." *Biometrics*(International Biometric Society) **29** (1): 159–183.

Sen, A. (1973). *On Economic Inequality.* Oxford, England: Clarendon Press.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), The social psychology of intergroup relations (pp. 33–47). Monterey, CA: Brooks/Cole.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behaviour. In S. Worchel & W. G. Austin (Eds.), *Psychology of Intergroup Relations* (pp. 7–24). Chicago, IL: Nelson-Hall.

Turner, John; Oakes, Penny (1986). "The significance of the social identity concept for social psychology with reference to individualism, interactionism and social influence." *British Journal of Social Psychology* **25** (3): 237–252.

Turner, J. C. (1999). "Some current issues in research on social identity and self-categorization thoeries." In Ellemers, N.; Spears, R.; Doosje, B. *Social identity* (Oxford: Blackwell): 6–34.

Wallman, S. (1977). "Ethnicity research in Britain," *Current Anthropology*, v. 18, n. 3, pp. 531–532.

Weiner, J. and Solbrig, O. T. (1984). "The Meaning and Measurement of Size Hierarchies in Plant Populations." *Oecologia* **61**, 334-336.

APPENDIX 1

Theoretical Value for Anderson-Darling Test for Normal Distribution

Case 1:

The mean μ is unknown and is estimated by $\bar{X}$, but population variance σ^2 is known. The Z is estimated by:

$$Z_1 = \frac{X_{(i)} - \bar{X}}{\sigma}$$

Case 2:

The mean μ is known, and the population variance σ^2 is estimated by $S^2 = \frac{1}{n}\sum_i \left(X_i - \bar{X}\right)^2$. The Z is estimated by:

$$Z_2 = \frac{X_{(i)} - \mu}{S_1}$$

Case 3:

Both parameters are unknown and are estimated by $\bar{X}$ and $S^2 = \frac{1}{n-1}\sum_i \left(X_i - \bar{X}\right)^2$

$$Z_3 = \frac{X_{(i)} - \bar{X}}{S}$$

Cases	Sample size	0.05	0.025	0.01
All cases	≥ 5	2.492	3.070	3.853
Case 1	≥ 5	1.087	1.285	1.551
Case 2	≥ 5	2.308	2.898	3.702
Case 3	≥ 5	0.752	0.875	1.035

Adapted from Tables 1, 2 & 3 *in* Stephen, M.A. (1979). "The Anderson-Darling Statistic." Technical Report No. 39, Oct. 31, 1979.

APPENDIX 2
Determining pValue

The observed T value could be converted to the Z score manually in three steps. Firstly, convert T to X by X = 1.15T. Secondly, insert X into the equation Z = (X – 1.64) + 1.98 – 0.004*n* – 0.06 where *n* is the sample size. Lastly, with known Z, obtained the pValue where *pValue = 1 – F(Z)*; thus:

$$F(Z) = \frac{1}{1 + \exp(-\sqrt{\pi}\left(\beta_1 Z^5 + \beta_2 Z^3 + \beta_3 Z\right)}$$

where β_1 = 0.0004406, β_2 = 0.0418198, and β_3 =0.90000000.

$Z = (X - 1.64) + 1.98 - 0.004N - 0.06$, and $X = 1.115T$.

OFFICE OF THE EDITOR-IN-CHIEF
International Journal of Research & Methodology in Social Science (IJRMSS)
420 Bourichane Rd., Unit 1 Ban Naxay
Xaysetha, Vientiane LAOS
Tel. +856 21 263585
Email1: Editor@socialsciencepublication.com
Email2: EditorIJRMSS@gmail.com
Website: www.socialsciencepublication.com

JOURNAL

The International Journal of Research and Methodology in Social Science (IJRMSS) is a double blind peer-review, open access, academic research, published quarterly. The journal publishes research articles, literature review articles, technical commentaries, and case studies.

MISSION

The mission of the journal is to provide publication service to the academic and research community in social science that is: (i) peer reviewed, (ii) high exposure, (iii) true open-access, (iv) fully achieved and retrievable online, and (v) leads to impact factor score.

OBJECTIVE

The objective of the journal is to provide an open access online forum and resource for researchers to share, disseminate, and exchange ideas and research outcome through peer reviewed publication.

POLICY

It is the policy of the journal to provide high quality research publication that is a true open access journal. The journal is committed to "no fees – no charge" to authors. All articles will be subjected to blind peer review process. We are committed to high quality research publication with free online access and strictly observe non-predatory publication policy.

PRINCIPLES OF TRANSPARENCY AND BEST PRACTICE IN SCHOLARLY PUBLISHING

The IJRMSS subscribes to the Principles of Transparency and Best Practice in Scholarly Publishing as recommended by Committee on Publication Ethics (COPE), the Directory of Open Access Journals (DOAJ), the Open Access Scholarly Publishers Association (OASPA), and the World Association of Medical Editors (WAME). A copy of the Principles of Transparency and Best Practice in Scholarly Publishing may be downloaded here: CLICK.

CALL FOR PAPERS

IJRMSS is published quarterly. Our call for papers is on-going. Authors may submit the research article(s) by email to: Editor@socialsciencepublication.com or EditorIJRMSS@gmail.com. We welcome articles in the following fields and their interdisciplinary

SOCIAL SCIENCE	**INTERDISCIPLINARY**	*Forestry*
Anthropology	*Applied physics*	*Library science*
Archaeology	*Artificial intelligence*	*Mathematical & Theoretical biology*
Criminology	*Bioethics*	*Mathematical physics*
Demography	*Bioinformatics*	*Medicine*
Economics	*Biomedical engineering*	*Military science*
Geography (human)	*Biostatistics*	*Network science*
History	*Cognitive science*	*Neural engineering*
International relations	*Complex systems*	*Neuroscience*
Law	*Computational linguistics*	*Science studies*
Linguistics	*Cultural studies*	*Scientific modeling*
Pedagogy	*Cybernetics*	*Semiotics*
Political science	*Environmental science*	*Sociobiology*
Psychology	*Environmental social science*	*Statistics*
Science education	*Environmental studies*	*Systems science*
Sociology	Ethnic studies	*Urban planning*
	Evolutionary psychology	*Web science*

All papers submitted must conform to the following format:

Font:	Times New Roman 12 points
Margin:	Single line spacing; 1 inch top, bottom, right & left.
Reference:	APA style. For example Click.
File:	MS Words. Originality will be checked by TURNITIN.
Template:	DOWNLOAD

ERRATA

In case where an error escapes the review process or is discovered after the publication has already been released, IJRMSS will contact the author for written comments and/or correction, and the journal will publish the correction in the *errata* section in the subsequent issue no later than 2 publication cycles after the error has been discover. Errata will be published when the error is material and may mislead or misrepresent facts. Minor error, such as typographical error or unless it involves mathematical formula, does not compel errata publication.

PLAGIARISM

IJRMSS maintains strict policy on originality of works and proper credit given to prior sources. We employ computer software to check for work originality. The article found more than 10% similarity with prior works will be automatically rejected for cause and no invitation for resubmission would be issued. Originality check does not count instances of citation and repeated citation of names, titles, and/or sentences and paragraphs in quotation or in-text quotation when proper quotation marks are used.

DISCLAIMER

The opinions expressed in each article are those of the individual author(s). Specific questions about the article should be directed to the corresponding author(s). The contact information of the corresponding author is listed in the footnote on the first page of the article marked with a star: ★. IJRMSS has editorial discretion to decline articles that are demeaning or disrespectful or otherwise intended to cause public outcry in matters of race, national origin, religion, ethnicity, and/or sexual orientation where public sensitivity outweighs the academic value of the proposed article(s).

INTERNATIONAL JOURNAL OF RESEARCH & METHODOLOGY IN SOCIAL SCIENCE

VOL. 1, NO. 2	APR. – JUN., 2015	QUARTERLY JOURNAL	ONLINE

Scientia,
Fortitudo &
Victoria

OPEN ACCESS

INTERNATIONAL JOURNAL OF RESEARCH & METHODOLOGY IN SOCIAL SCIENCE
Online Publication

OFFICE OF THE EDITOR-IN-CHIEF
Email: Editor@socialsciencepublication.com

Registered Office	Correspondence in Thailand
420 Bourichane Rd., Unit 1 Ban Naxay	55 / 123 Rangsit-Nakhon Nayok Rd.
Xayasetha, Vientiane, LAOS	T. Lampakkood, A. Thanyaburi
Tel. +856 21 263585	Pathumtanee, THAILAND
Email: EditorIJRMSS@gmail.com	Email: EditorIJRMSS@gmail.com

www.socialsciencepublication.com