

Quimba, Francis Mark A.; Barral, Mark Anthony A.

Working Paper

Exploring neural network models in understanding bilateral trade in APEC: A review of history and concepts

PIDS Discussion Paper Series, No. 2018-33

Provided in Cooperation with:

Philippine Institute for Development Studies (PIDS), Philippines

Suggested Citation: Quimba, Francis Mark A.; Barral, Mark Anthony A. (2018) : Exploring neural network models in understanding bilateral trade in APEC: A review of history and concepts, PIDS Discussion Paper Series, No. 2018-33, Philippine Institute for Development Studies (PIDS), Quezon City

This Version is available at:

<https://hdl.handle.net/10419/211053>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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

DISCUSSION PAPER SERIES NO. 2018-33

Exploring Neural Network Models in Understanding Bilateral Trade in APEC: A Review of History and Concepts

Francis Mark A. Quimba and Mark Anthony A. Barral



The PIDS Discussion Paper Series constitutes studies that are preliminary and subject to further revisions. They are being circulated in a limited number of copies only for purposes of soliciting comments and suggestions for further refinements. The studies under the Series are unedited and unreviewed. The views and opinions expressed are those of the author(s) and do not necessarily reflect those of the Institute. Not for quotation without permission from the author(s) and the Institute.

CONTACT US:

RESEARCH INFORMATION DEPARTMENT
Philippine Institute for Development Studies

18th Floor, Three Cyberpod Centris - North Tower
EDSA corner Quezon Avenue, Quezon City, Philippines

publications@mail.pids.gov.ph
(+632) 372-1291/(+632) 372-1292

<https://www.pids.gov.ph>

Exploring Neural Network Models in Understanding Bilateral Trade in APEC: A Review of History and Concepts

Francis Mark A. Quimba
Mark Anthony A. Barral

PHILIPPINE INSTITUTE FOR DEVELOPMENT STUDIES

December 2018

Abstract

Machine Learning algorithms, such as the artificial neural networks, are acknowledged to outperform several econometric procedures in predictions. Machine learning becomes popular for doing operations that practically require more efficiency and accuracy, derived basically from the algorithm's powerful learning capability that can be applied in many fields, particularly in anticipating future event.

Decision- and policy-makers have always been concerned in predicting the future in order to implement appropriate decision and policies. This necessitates the need to improve the accuracy of prediction and utilize more superior algorithms.

This paper seeks to understand certain frameworks that can be used to improve the analysis and prediction of trade flows in APEC using neural networks. Discussions include the history of neural network development, the biological neuron, the artificial neuron, and the potential use of neural networks in trade analysis. This paper also compares the different estimation procedures of the gravity model – OLS, PPML, and GPML – with the neural network. The neural network estimation of the gravity equation is found superior over the other procedures in terms of explaining the variability of the dependent variable (export) around its mean, and in terms of the accuracy of predictions.

Keywords: neural networks, artificial intelligence, machine learning, gravity model, OLS, PPML, GPML, trade, APEC

Table of Contents

1. Introduction	3
2. Review of Literature	4
2.1. The Birth of Neural Network.....	4
2.1.1. First Phase	4
2.1.2. Second Phase	5
2.1.3. Third Phase.....	6
2.1.4. Fourth Phase	6
2.2. AI, Machine Learning and Neural Networks	6
2.3. Neural Network for Economic and Policy Analysis	8
3. Neural Network Application	10
3.1. The Artificial Neural Process.....	10
3.1. Application to APEC Bilateral Trades.....	12
3.1.1. Data and Methodology	12
3.1.2. Results of Estimation Procedures	13
4. Conclusions/Recommendations	16
5. Bibliography.....	17

List of Tables

Table 1 presents some of the theories and assumptions of the estimators.	13
Table 2. Results of Gravity Model Estimation	13
Table 3. Comparison of Model Accuracy	15

List of Figures

Figure 1. Biological neural network.....	7
Figure 2. The Neural Process (Castrounis 2018).....	10
Figure 3. Artificial Neural Network	11
Figure 4. Sigmoid Function.....	11
Figure 5. Estimated Neural Network	15

Exploring neural network models in understanding bilateral trade in APEC: A review of history and concepts

Francis Mark A. Quimba and Mark Anthony A. Barral*

1. Introduction

Bilateral trade flow represents the value of goods and services that a country exports. It is an important indicator used by policymakers in understanding how international trade policy and domestic policies interact. Understanding bilateral trade flow is necessary as net export is an important determinant of a country's economic development and growth.

In doing trade policy research, there is a wide array of conventional statistical and econometric procedures one can use and exploit. In conjunction, there is also a growing recognition for machine learning algorithms, such as the neural networks, that are seen to have high superiority over traditional econometric models. The neural network algorithm is acknowledged to approximate any function arbitrarily well and can be used to achieve superior bilateral trade flow estimation or prediction.

In addition, the data revolution, resulted from modern and more efficient techniques of data gathering and processing, contributes to the scope and quality of results of statistical analysis and methods. Large datasets allow better measurements of economic effects and outcomes, and can help pose new sorts of questions and enable novel research designs that can inform the consequences of economic policies and events (Einav and Levin 2013). So when it comes to big datasets, different tools may be required to address issues unique to them. For instance, there may be more potential predictors appropriate for estimation, which require some kind of variable selection. Also, large datasets may allow for more complex relationships than simple linear, which machine learning techniques or algorithms, such as decision trees, support vector machines, neural nets or deep learning may model (Varian 2014).

This paper is motivated by the availability of modern techniques, such as neural network algorithm, and their potential uses in the field of economic, while being prompted by the challenge of exercising more precise decision- and policy-making. This study aims to determine the applicability of neural network particularly in understanding bilateral trade within APEC. This study also recognizes the necessity in related and other areas of study to develop techniques or improve existing ones to address problems that are complicated and are currently not well understood. The succeeding sections provide historical accounts on the development of theories and concepts that founded the development of principles of artificial intelligence, machine learning and neural networks. The applications of neural networks in economic is also discussed, followed by an actual example of application of neural network to APEC bilateral trade, in comparison with the gravity model estimation using OLS, PPML (Poisson Pseudo Maximum Likelihood), and GPML (Gamma Pseudo Maximum Likelihood).

* Research Fellow and Research Associate, Philippine Institute for Development Studies

2. Review of Literature

2.1. *The Birth of Neural Network*

The origin, development and evolution of neural networks are associated with the development of artificial intelligence, which stemmed from the interests, ideas, and imaginings of man in his desire for answers to his queries about life and the absence or end of it. Man's desire led his curiosity to emulate or recreate life itself, or perhaps parcels of it, as evidently manifested in the Greeks' mechanical man and artificial beings, such as Hephaestus' mechanical servants that, with the bronze man, incorporated the idea of intelligent robots. Also, myths involving human-like artifacts and mechanical toys that were actually constructed by, e.g., Archytas of Tarentum, Hero, Daedalus and other real persons (AITopics n.d.). Thus, the evolution of man's desire to understanding life led to the simultaneous developments of theories of immaterial mind, translation or representation of ideas, and mechanical simulation of cognitive processes. The history and development of neural network can basically be divided into four phases. The first phase considers the birth and evolution of intangible ideas and thoughts before it transforms into the tangible mechanics and digits. The second phase is about understanding the logic behind the behavior of biological neuron and its ability to learn. The third pertains to the diminished interests in the field of neural network. The fourth stage is the renewed interests in neural network until the recent history.

2.1.1. First Phase

The initial stage of the neural network development is considered to have started from the ancient men's curiosity and fascination about life, about man himself, how life happens and what keeps man living.

Ancient Greece associated life with the last breath (breeze or wind) that comes out of the mouth during death, which they consider to be the "soul" that keeps man alive (Frede and Reis 2009). Later, Aristotle theorized how this soul is able to move the body, which came to mean his metaphysics that attributed the movement into three divisions of soul – the liver for appetite, heart for the senses, and brain for the rational parts. He regarded the heart as the controller of both voluntary and involuntary movement, but no regard to the brain as the overall source of animal behavior (Bishop 2014). By 4th Century BC, he invented the Syllogistic Logic, the first formal deductive reasoning system, which later on became dominant in Western thought (AITopics n.d.). Aristotle's theories about living creature and reasons were later adopted or developed by various scholars, such as Thomas Hobbes (1651) (Dalakov 2018) and Descartes (1649) who laid the foundations of a dualist view of the original concepts of Aristotle on soul. According to Decartes, the immaterial soul (mind) (*res cogitans*) interacts with the material body (*res extensa*) at the pineal gland. This view of the brain as the overall controller of behavior was later empirically supported by Thomas Willis, and English physician, in 1667, who identified links between the physical structure and the pathological behaviors (Bishop 2014).

This fascination of man was passed on to generations and was made evident by his endeavor to imitate life through his invention of machines that mimic the behavior of man, such as mechanical animals, novelties, and other machines that also paved the way to the development of machines that can perform calculations (Skirry, n.d.; Oppy and Dowe, 2008).

In 1642, Blaise Pascal built the first considered digital calculating machine that paved the development of modern computers (Educalc.net 2018). 1674, Gottfried Wilhelm von Leibniz, a German polymath, created the first mechanical calculator, described as a machine for solving algebraic equations. In 1675, he made a comparison between reasoning and mechanism to reduce reasoning to a kind of calculation and to build a machine that is capable of performing such calculations. Later on, Leibniz adopted the binary logic system, which developed by Pingala in 300 BC, and learned its importance in mathematical operations (Dalakov 2018).

In 1854, George Boole published “An Investigation of the Laws of Thought”. Boole sought to stretch the boundaries of syllogistic logic to represent and manipulate all logically valid inferences using a general method (Barnett 2011). His system of symbolic logic is a form of algebraic logic based on truth values true and false (1 and 0) and the conjunction (AND), disjunction (OR), and negation (NOT) operators. This laid the foundation for what is now known as the Boolean algebra (Johnson, 2016; and Lande, 2014), which is the mathematical foundation of digital circuits, as it specifies the relationship between Boolean variables used to design combinational logic circuits using logic gates (Emant Pte Ltd. 2018). The Boolean logic of reducing thought to 1s and 0s became the foundation for all digital computing (Chisling n.d). It was adopted as a theory of mathematics in which all variables are either “true” or “false”, or “on” or “off”, and was applied to electrical circuits by Claude Shannon in the 1930s that eventually led to the development of modern computers (Titcomb 2015). Until several years after, in 1936, Alan Turing, an English Mathematician, launched computer science as a modern intellectual discipline, and proposed a model for what is now known as the Turing machine, a universal machine capable of computing any computable by any other Turing machine (Erickson 2014).

2.1.2. Second Phase

The second phase pertains to the discovery of the behavior of neuron that controls mental functions and dictates the behavior of man. The mechanism of biological neuron inspired many enthusiasts to develop artificial systems that resembles how the neuron functions.

The form of logic and mathematical notion of computation developed based on the Turing machine were employed by Warren S. McCulloch and Walter H. Pitts in their 1943 paper, “A logical Calculus of the Ideas Immanent in Nervous Activity”, to explain how neural mechanism understands mental functions (Piccinini 2014). In their model of artificial neural networks, they postulated that the neuron can be in a binary state, either on or off. The model was equivalent to the Turing machine and they proved that some network of connected neurons could learn and compute any computable function. This was considered as the first work recognized in the field of artificial intelligence (Negnevitsky 2005).

Several years after, in 1948, Shannon published his “A Mathematical Theory of Communication”, where he introduced fundamental concepts about information as “uncertain” and is “digitally transmitted”. He modeled information source as superfluous and continuously transmitted as a random variable or random process (Yeung 2001). He also defined entropy as the only measure of information that is (a) continuous in the probabilities, (b) increasing with n-values if the random variable is equiprobable, and (c) additive, thus if the random variable is the result of two choices, entropy can be obtained by summing the entropies of the two choices (Verdu 1998). He then employed probability to develop the theory of information and theorized

that information can be converted into a stream of 0s and 1s called bits that can be delivered to the receiver correctly without any reference to their actual meaning. It is in this work of Shannon where the term bit was used (binary digit) (Yeung 2001). Because of his works, Shannon was named “father of the digital age”, “father of information theory, or “father of modern digital communications and information theory” (Aftab et al. 2001 and Geoghegan 2008).

Influenced by McCulloch and Pitts, Marvin Minsky started an attempt to construct a working artificial neural network device that he named SNARC (Stochastic Neuro-Analog Reinforcement Computer). It was a hardwired electro-technical device made of several identical artificial neurons connected in a network and was capable of learning simple concepts (Klemen 2007). After a year, in 1951, together with Dean Edmons, he finally completed the first neural network computer. They were encouraged and supported by von Neumann (Negnevitsky 2005).

While Minsky was still completing his machine, in 1950, Turing published a paper, “Computing Machinery and Intelligence”, which led to the development of another field of science that would be called “Artificial Intelligence”, as coined by another scientist, John McCarthy, years after. Turing started his seminal paper by asking “Can machines think?”, and proceeded to propose a decision criterion, the Turing Test, for evaluating whether machines can actually think (Smith et al. 2006).

2.1.3. Third Phase

Throughout the modern history, several other inventions and innovations paved the way to the development of artificial intelligence, machine learning, and deep learning. From 1974 to 1980s, declining interests on artificial intelligence and machine learning was observed. The funding supports for neural network and artificial intelligence projects (AI winter) declined. Numerous enthusiasts, however, continued to conduct research and application of neural networks in various fields of expertise. In 1965, the first expert system, which automated the decision-making process and problem-solving behavior of organic chemists, was constructed. The 1970s was considered a period of knowledge-based approach (aside from AI) (Press 2016).

2.1.4. Fourth Phase

By 1980s, expert systems were popularized. It regained the interests on neural networks. Several years after, however, expert systems declined and the second AI winter was witnessed. By 1988, Judea Pearl published the “Probabilistic Reasoning in Intelligent Systems”, which is a “representational and computational foundation for the processing of information under uncertainty”. Judea was credited for the invention of Bayesian networks and the principal algorithms used for inference in these models. His works revolutionized the field of artificial intelligence and became important tool in engineering and natural sciences (Press 2016).

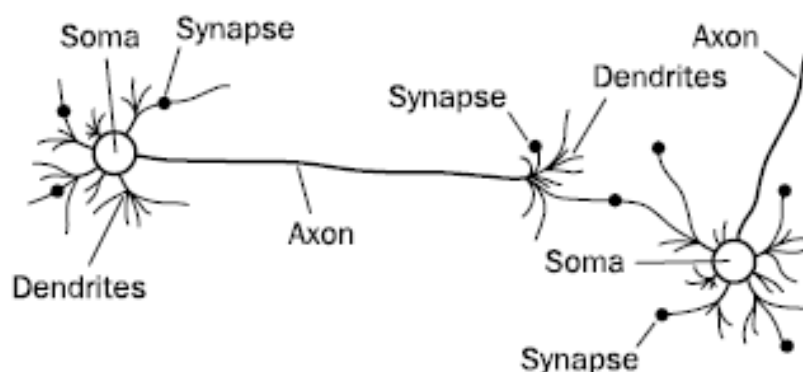
2.2. *AI, Machine Learning and Neural Networks*

AI is the theory and development of computer systems that are able to performs tasks that require human intelligence (Ganguly 2018). This concept of having machines with human senses is referred to as the “General AI”, which today remains in movies and science fiction

novels. “Narrow AI”, on the other hand, is the concept of having technology perform specific tasks, such as image classification and face recognition, and is what exists today. AI is achieved through machine teaching or the practice of using algorithms to parse, learn from, and make prediction out of a set of data (Copeland 2016). Machine learning is a subset of AI in the field of computer science and uses statistical techniques to give computer the ability to “learn” with data, without being explicitly programmed. Arthur Samuel coined the phrase “machine learning” in 1959 as “the ability to learn without being explicitly programmed (Ganguly 2018 and Colagrossi 2018).” It is the idea of writing a program that can create better programs that people can write directly using data (experience) (Littman 2006). Other names include classification learning, statistical learning, data mining, concept learning, discriminative learning, logistic regression, and pattern recognition (Littman, 2016). Machine learning involves adaptive mechanisms that enable computer to learn from experience, example, and analogy, which can improve the performance of an intelligent system over time. These mechanisms or algorithms form the basis for adaptive systems (Negnevitsky 2005). Algorithms, therefore, are used to model a problem based on its interaction with the experience or environment or whatever we want to call the input data. They are grouped according to their functions or how they work (Brownlee, 2013). Algorithms also include decision tree learning, inductive logic programming, clustering, reinforcement learning, and Bayesian networks, boosting, nearest neighbors, support vector machines (Copeland 2016 and Littman 2006). Artificial neural networks and genetic algorithms are the most popular algorithms (Negnevitsky 2005).

Artificial neural networks (also ANNs or neural nets) are analogous to non-parametric and non-linear statistical regression models. They can model non-linear processes with few or no assumptions about the character of the system being modelled, making them useful in business and economics where little is known about the nature and dependencies occurring in analyzed processes (Morajda 2003). A neural network is a model of reasoning based on the human brain, which is consisted of neurons, a set of interconnected nerve cells that serve as the basic information-processing units. There are nearly 10 billion neurons and 60 trillion connections (synapses) between them. A neuron is consisted of a body (soma), a fibers (dendrites), and a single long fiber (axon) that connects to dendrites and somas of other neurons (Figure 1) (Negnevitsky 2005).

Figure 1. Biological neural network



Majority of neurons encode their outputs or activations as series of electrical pulses (spikes, signals, action potentials, or information) that are propagated or transmitted along the axon to

another neurons in the network through a complex chemical reactions. Synapses weigh the transmission signals using neurotransmitters between axons and dendrites to build up large neural networks. Signals, in the form of chemical substances, are released from the synapses and are received by dendrites. These change the electrical potential of the cell body and when the potential reaches its threshold, the soma converts it into output activations. These electrical pulse or action potentials are then sent down again through the axon, and the pulse spreads out and eventually again reaches synapses, causing them to increase or decrease their potential. The synapse, therefore, can either be excitatory or inhibitory and weigh or transform the axon signal. If it is excitatory, the receiving neuron is also activated, whereas, if it is inhibitory, the receiving neuron remains inactive. Neuron switches from an inactive to an active state, thus the neuron is firing, and this happens in every few milliseconds and the cell continuously calculates its state. In response, the strength of neurons' connections change, new connections can be formed, and an entire collection of neurons may even migrate to another place. This characteristic of neuron is referred to as plasticity, and is the basis for learning in the brain (Negnevitsky 2005).

Hebb (1949), stated that “when an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased” (Wang and Raj 2017).

The ease and naturalness of biological neural networks to learn led to attempts to emulate this in a computer. Information is stored and processed in a neural network simultaneously throughout the whole network rather than at specific location, which means both data and its processing are global rather than local (Negnevitsky 2005).

Data mining or data science is a related and complementary field, which is a non-trivial process of identifying valid, novel, potentially useful, and untimely understandable patterns, structures, models, trends, or relationships in data to enable data-driven decision making. While statistics is the science of learning from data, and measuring, controlling and communicating uncertainty, data mining is not only modeling and prediction, but also an iterative process of problem solving that must be mastered through interdisciplinary and transdisciplinary effort. Traditionally, statistics is concerned with analyzing primary data and checking the validity of a specific existing idea; thus top-down (explanatory or confirmatory) analysis, or idea evaluation or testing. Data mining uses secondary data that have been collected for other reasons (not under the control of the investigator), and uses these data to create new ideas (hypotheses); thus bottom-up (exploratory and predictive) analysis or idea generation (knowledge discovery) (Statoo Consulting 2016).

2.3. Neural Network for Economic and Policy Analysis

Inspire from the ability of biological neuron with the ability to learn based on infinite number of observations, the neural network provides superiority in data analysis. Neural network represents an agent who learns from his environment and, therefore, develops strategies based on what has been observed. This superiority of the neural network makes it an excellent choice for solving real economic issues and allows it to capture relationships among different factors and use this in the assessment of potential risks, which is important in making decisions, particularly on finance, trade and other fields.

Neural network can be efficient in solving macroeconomic issues. Villa and Valaitis (2018) developed a global solution method to solve large state space macro-finance models using artificial neural network expectation algorithm, which they found out to be faster, precise, and more scalable than the standard parameterized expectation algorithm. This, as they suggested, could allow to study more easily the optimal policy with multiple maturities.

Similarly, Anastasopoulos et al. (2017), provided a framework for political image analysis using deep neural network. They utilized developments in machine learning to use images to gain systematic knowledge about how politicians and political organizations interact with the voters. Images of politicians can manipulate the opinion and sentiment of the people. Developments in computer vision allows for accurate and faster identification of labeled data, to provide broad framework for political image analysis.

Moreover, neural networks, among other algorithms, also provide for an opportunity to develop procedures in benefiting from the use of big data, especially in detecting patterns and behaviors that are not usually identified using the conventional means.

Decades ago, data on economic activity was relatively scarce but this has dramatically changed as technology developed, especially with the advent of internet where everything is recorded. These records include queries and clicks on search, purchase information online purchase platforms, online news, videos, and any activity done that provide information about a person's behavior. A person's behavior is also recorded in text messaging, cell phones, geo-locations, scanner data, employment records, among others, which are all part of data footprint. Data now-a-days is generated faster, has greater coverage and scope, and includes new types of observations and measurements that previously were not available (Einav and Levin 2013). Big data is used for tracking business processes and outcomes to improve an organization's efficiency and for predictive modeling to generate information. For example, Amazon and Netflix recommendations are based on predictive models of individuals' book or movie preferences; Google's search results and news feeds rely on algorithms that predict the relevance of web pages or articles; Apple's auto-complete function predicts a text based on the usage patterns; and online advertising and marketing target potential customers based on automated predictive models (Einav and Levin 2013).

Predictive algorithms are not limited to online and internet application. In health care, insurance providers adjust payments and quality measures based on "risk scores" derived by predicting a person's health costs and outcomes. Health score is typically a weighted sum of health indicators that identify chronic conditions. The weights are based on statistical analysis. Predictive algorithms are also used by credit providers and by authorities in identification of terrorist threats, among others. These applications rely on converting large amounts of unstructured data into predictive scores, often fully automated and scalable, and sometimes real time (Einav and Levin 2013).

Similarly, the government collects enormous amounts of granular administrative data, such as those from social security and internal revenue, which are most-likely underutilized, both by government agencies and by researchers and private data vendors who have limited and restrictive access. These large amount of data can be used to inform policy decision and improve government services. The use of government administrative datasets often has profound consequences for economic policy discussion, even from simply describing basic patterns in the data. For example, using Internal Revenue Service (IRS) data, Piketty and Saez,

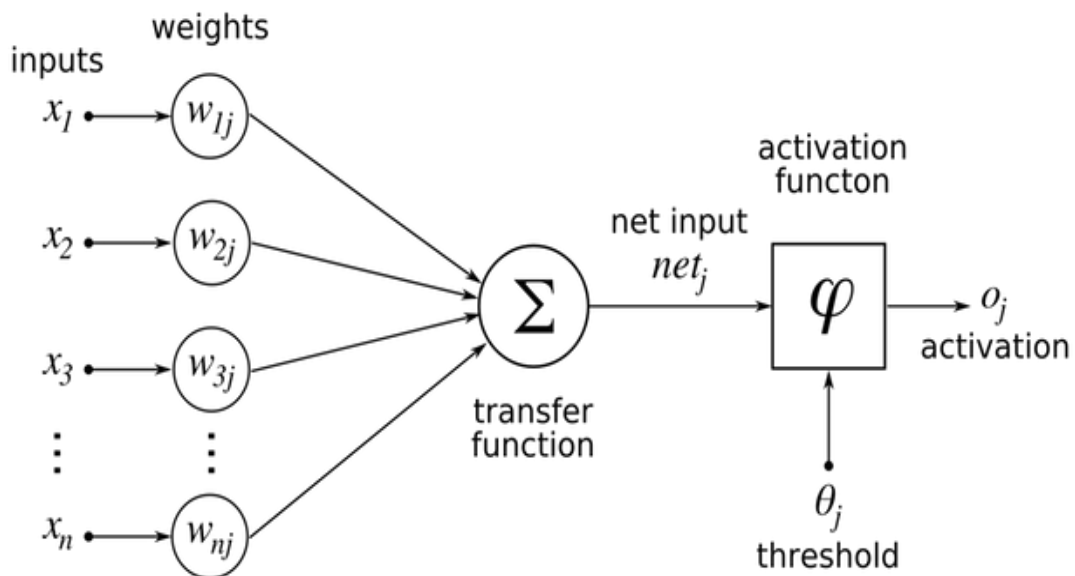
in 2003, derived a historical series of income shares for the top percentiles of earners among US households. Their paper and related works influenced the recent policy debates by helping make the rising income share a major focus of discussions about economic inequality. Similarly, John Wennberg and his colleagues at Dartmouth used large samples of Medicare claims to show a great variation in Medicare spending per enrollee that cannot be attributed to the differences in health status or prices and does not correlate with measured health outcomes. This research has become the leading evidence for inefficiency in the US healthcare system during the debate on Affordable Care Act in 2009. Making large datasets accessible, and using them, could be an opportunity for the government to improve the functions of agencies (Einav and Levin 2013).

3. Neural Network Application

3.1. The Artificial Neural Process

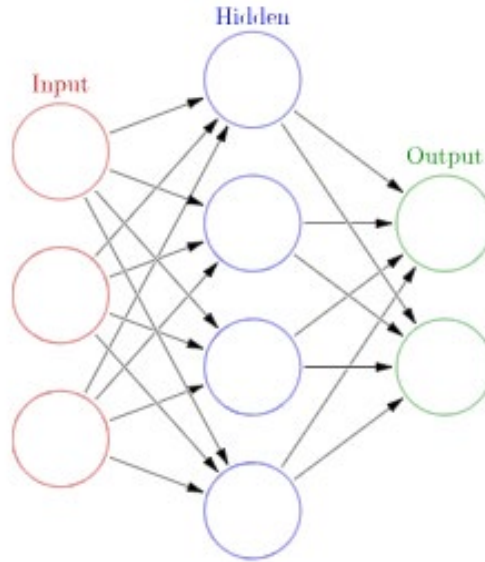
Artificial neural network is a model of the biological neural network, based on the assumptions on how the biological neuron behaves from different stimuli or other factors. Basically, a single neuron is able to transmits message to another neuron if the sum of the weighted input signals exceeds a certain threshold level. This is called the activation process. The input signal a neuron transmitted is a weighted combination of several other inputs that may have different influence. These weighted inputs are applied with a certain transfer function that can be transformed mathematically prior to the evaluation if the activation threshold has been exceeded. This can be described linearly or nonlinearly (Castrounis 2018).

Figure 2. The Neural Process (Castrounis 2018)



The ANN structure is inspired by the biological neural networks. The basic architecture of ANN consists of input layer, hidden layer and output layer. This can be developed further by increasing the number of hidden layers and even the output layer (Figure 3) (Castrounis 2018).

Figure 3. Artificial Neural Network



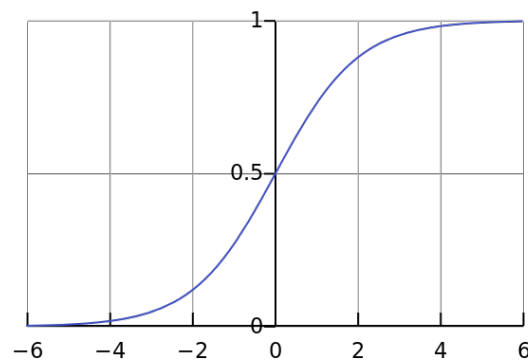
The output neuron can be calculated as

$$a = f\left(\sum_{i=0}^N w_i x_i\right)$$

where, f is the activation function.

There are a number of activation functions that can be used. One commonly used activation function is the sigmoid function (Figure 4), which takes a value between 0 and 1 by applying a threshold (Maladcar 2018).

Figure 4. Sigmoid Function



3.1. Application to APEC Bilateral Trades

In investigating the potential use of neural network models to understand the bilateral trade among APEC economies, this paper employs the gravity model to be estimated using OLS, PPML and GPML. OLS is used as the base model while PPML and GPML are selected as these are suggested in the literature to yield relatively better estimates of gravity model.

The PPML (Poisson Pseudo Maximum Likelihood) and the GPML (Gamma Pseudo Maximum Likelihood) are estimation methods for gravity models belonging to the general linear model. The PPML is estimated using the quasi-poisson distribution and a log-link, while GPML is estimated using the gamma distribution and a log-link. Both the estimation procedures can be used for both cross-sectional and panel data (Woelwer, et. al, 2018).

The gravity model is considered the workhorse tool in analyzing international trade. It measures the relationship between the trading behavior of two countries based on their GDPs and distances. It provides accurate predictions of international trade and has strong theoretical foundations (Wohl and Kennedy 2018).

The performances of gravity model estimations are compared to the performance of neural networks using the MSE (mean squared error) and RMSE (root mean squared error) (other metrics may also be used). MSE is a squared RMSE, an absolute measure of fit and is reported in the same unit as the response variable.

3.1.1. Data and Methodology

The data used include exports data (USD millions) GDP (current prices in USD millions) and population (in thousands) from the World Bank, geographic (distance, contiguity, area, etc.) and cultural/historical data (language, colony, etc.) from CEPII. The data covers 21 years from 1996 to 2001 for all 21 economies in APEC, with 8,329 observations.

The models are compared based on their predictive capacity. In order to do this, the dataset was divided into two sets, the training observations (1996 – 2013) and the test observations (2014 – 2017). Training set is 81 percent of the dataset, while test set is 19 percent.

The gravity equation for the bilateral trade used in the study used the conventional information used in gravity model literature, such as value of exports, distance, income levels, population, area, common language, colony, and contiguity. This study also considers whether the reporting and partner economies have FTA, as well as the their membership status in ASEAN, ASEAN+3, and OECD.

In OLS estimation, the continuous variables (distance, GDP, etc.) and the dependent variable are transformed into logs, while in the PPML and GPML, the natural logs of the continuous variables are used and the export values are untransformed. The neural network, on the other hand, does not require the data to be transformed into logarithmic form, however, the variables are scaled so that their means equal zero and the standard deviations equal one.

Table 1 presents some of the theories and assumptions of the estimators.

Estimator	Foundational Theory
OLS	Assumes the variance of the error is constant across observations (homoscedasticity); in the presence of heteroscedasticity, however, nonlinear estimation issued as the OLS estimation may yield inconsistent estimates; drops the zeros from the sample when trade values transformed into log form
PPML	Reduces the magnitude of the coefficients and the standard errors; assigns the same weight to all observations; preferred estimation in the presence of heteroscedasticity as it does not require any further information on the pattern of heteroscedasticity; performs very well with large proportion of zero
GPML	Similar to PPML but the conditional variance of the dependent variable is assumed to be proportional to its conditional mean; it assigns less weight to observations with a larger conditional mean; estimates are found to be adequate in the presence of heteroscedasticity
NN	Based almost completely on heuristics that means there has no theory about the selection of its architecture; contemporary neural network is based on the evolutionary physics from which the idea to train and capture architectural heuristics about the network transpired

Sources: Wohl and Kennedy 2018, Yotov, et al. 2016, Yotov and Piermartini 2016, Marin 2015, Lipton 2015, Gomez-Herrera 2013, and Tillema, et al. 2006

3.1.2. Results of Estimation Procedures

The results of the OLS, PPML, and GPML estimation of the gravity model return almost the same results with respect to the significance of the variables. Distance, income levels, language, colony, and FTA (whether reporting and partner economies have FTA), are consistently significant at 1% level in all estimation procedures. The significance of other variables, however, are reduced to 5% and 10% levels. These include the ASEAN partner (status of partner economy's membership to ASEAN), OECD reporter (status of reporter economy's membership to OECD) (Table 2).

Table 2. Results of Gravity Model Estimation

Variables	Coefficients		
	OLS	PPML	GPML
(Intercept)	1.2377	5.0713	7.6360
Distance	-0.3737***	-0.5899***	-0.9700***
GDP reporter	0.5099***	0.6441***	0.7244***
GDP partner	0.4334***	0.6772***	0.6870***
Population reporter	0.0028***	0.2848	0.0885
Population partner	0.0200***	0.2910	0.3607
Area reporter	-0.0070***	-0.2205	-0.1390

Area partner	-0.0115***	-0.2080	-0.2153
Language	0.0289***	0.4544***	0.4959***
Colony	0.0179***	0.1738***	0.4115***
Contiguity	0.0024**	0.4520***	-0.0806***
FTA	0.0179***	0.4155***	0.4092***
ASEAN reporter	0.0007**	-0.0519***	0.0080***
ASEAN partner	0.0183***	0.3783*	0.1886*
ASEAN+3 reporter	0.0284***	-0.0237***	0.4356***
ASEAN+3 partner	-0.0070	-0.5838**	-0.1520**
OECD reporter	0.0047***	-0.1220*	0.1708*
OECD partner	0.0294***	0.0283	0.4801
r-squared	0.7013	0.3304	0.3304
Adj. r-squared	0.7005	0.3287	0.3287

*** Significant at 1%

**Significant at 5%

*Significant at 10%

Population and area of the reporting and partner economies, as well as the status of OECD membership of partner economy, appear to be insignificant in PPML and GPML. On the other hand, the significance of contiguity improves from 5% in OLS to 1% level of significance in PPML and GPML, while ASEAN membership status of partner economy becomes significant in PPML and GPML.

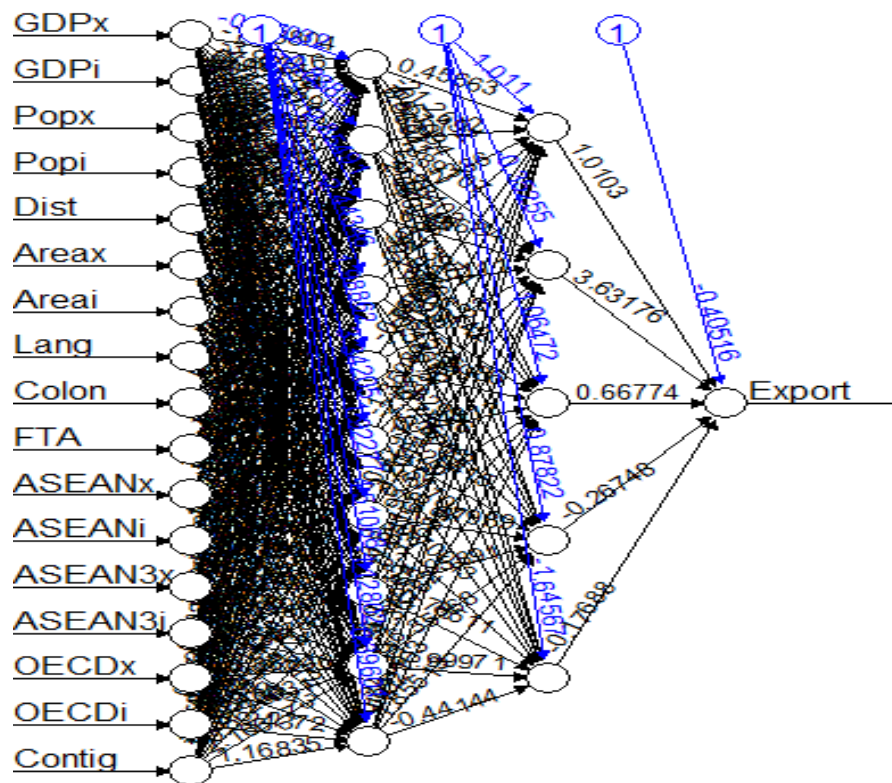
It is interesting to discover that the influence of contiguity is reversed in the GPML. This may indicate that contiguity does not necessarily mean strong trading relationship. But more into these must be looked upon deeper. Two contiguous economies, perhaps, differ in structure, politics and capacities to make significant trade.

Similar to contiguity, the direction of influence of reporter economy's membership status to ASEAN and OECD is also reversed in PPML.

It is important to note, however, that the PPML and GPML are only able to explain about 33 percent of the variability of exports around its mean, as compared to OLS that is able to explain 70 percent of the variability.

Using the same datasets and variables, the neural network model was estimated. Two hidden layers are randomly used in constructing the neural network, with the first hidden layer having ten nodes, while the second hidden layer with five nodes. The blue lines represent the biases while the black lines are the coefficients (Figure 5).

Figure 5. Estimated Neural Network



The figure presents the plotted neural network. The black lines, with the associated numbers, represent the weights that indicate how much each variable contributes to the node. The blue lines are the bias weights, which allows the activation function to shift (i.e. change in steepness of the sigmoid).

As there is currently no standard procedure to determine the appropriate number of layers or nodes that can be used, the numbers of hidden layers and nodes used in the study are randomly selected. It is noted, however, that adding units can improve the performance of the neural network on the training data but can also increase the errors in test data as the network becomes less generalized (Wohl and Kennedy 2018).

Comparing the different estimation procedures, it is remarkable to learn that PPML and GPML have much lower MSE and RMSE than the OLS estimator. However, the results for the neural network show that the algorithm is more superior in estimating the influences of the variables to the export performances within APEC (Table 3).

Table 3. Comparison of Model Accuracy

	r-squared	Adj. r-squared	Out-of-sample MSE	Out-of-sample RMSE
OLS	0.7013	0.7005	313.5784	17.7081
PPML	0.3304	0.3287	24.9394	4.9939
GPML	0.3304	0.3287	51.9491	7.2076
Neural Network	0.7925	-	0.0012	0.0351

The RMSE as standard deviation of the residuals (prediction errors) tells that while OLS has out-of-sample RMSE or deviation from the actual exports data by US\$17.7 million. While PPML and GPML are able to closely predict the actual export values with much lesser deviation of US\$5.0 million and US\$7.2 million respectively. Neural network, on the other hand, is able to almost perfectly estimate the actual values with a deviation of US\$0.6 million. To add, the neural network is able to explain 79 percent variability of exports around its mean. Both the RMSE and r-square values are able to indicate the superiority of neural network compared to OLS, PPML and GPML.

These results support the findings of Maliki et al. 2014, Bajracharya 2010, and Chen 2005, among others, that described the predictive superiority of the neural network model.

4. Conclusions/Recommendations

The study compares the analysis of gravity models (OLS, PPML, and GPML) and neural network using panel data in analyzing the bilateral trade flows between APEC economies.

In the literature, it is widely accepted that the main limitation of a neural network model is that it fails to describe causal relationship between the variables. Thus, it is currently not possible to select significant policy variables among the set of hypothesized variables. It is, however, possible to combine the two procedures, regression model and neural network. Regression model is used to select significant variables and the superiority of neural network in predicting can be applied into it for forecasting bilateral trades. The need for further research must be recognized particularly on interpreting the estimates of the neural network.

The paper reveals that a theoretically sound model to design policies can be backed-up by a model with superior predictive capacity. The results of the study revealed the capacity and potential of neural network in making more accurate estimation and prediction, especially in understanding bilateral trade. The accuracy of neural network can be very beneficial for policymakers and researchers.

For research directions, the superiority of neural network can potentially aid in understanding, for instance, the interregional freight distribution, predict the effects of trade agreements or other policies, and can even be applied to specific trade commodities and sectors.

Moreover, it is also important to understand how and what factors can affect the accuracy of the neural network or how the model can be improved. Perhaps, a way to determine a sufficient and efficient number of nodes and layers that must be used in neural network modeling can be designed.

5. Bibliography

- Anastasopoulos, J., D. Badani, C. Lee, S. Ginosar, and J. R. Williams. 2017. Political Image Analysis with Deep Neural Networks.
- Aftab, O., P. Cheung, A. Kim, S. Thakknar, and N. Yeddnapudi. 2001. Information Theory: Information Theory and the Digital Age. 6.933 Project History, Massachusetts Institute of Technology. (Retrieved from <http://web.mit.edu/6.933/www/Fall2001/Shannon2.pdf> on September 19 2018).
- AITopics. n.d. A Brief History of AI. (Retrieved from <https://aitopics.org/misc/brief-history> on September 17 2018).
- Bajracharya, D. 2010. Econometric Modeling Vs Artificial Neural Networks – A Sales Forecasting Comparison. University of Boras Master's Thesis. (Retrieved through <http://bada.hb.se/bitstream/2320/7986/1/2010MI17.pdf> in December 2018).
- Barnett, J. H. 2011. Origins of Boolean Algebra in the Logic of Classes: George Boole, John Venn, and C.S. Pierce.
- Bishop, J. M. 2014. History and Philosophy of Neural Networks. (Retrieved from https://pdfs.semanticscholar.org/a481/c9f2f6052ef900558f4da307df4bee075524.pdf?_ga=2.33578722.215681743.1537406938-1961704785.1531893570 on September 20 2018).
- Castrounis, A. 2018. Artificial Intelligence, Deep Learning, and Neural Networks, Explained. (Retrieved from <https://www.kdnuggets.com/2016/10/artificial-intelligence-deep-learning-neural-networks-explained.html> in December 2018).
- Chen, J. 2005. Neural Network Applications in Agricultural Economics. University of Kentucky Doctoral Dissertations. 228. (Retrieved from https://uknowledge.uky.edu/gradschool_diss/228/ in December 2018).
- Chisling, A. n.d. The History of AI is a Neural Network of the Greatest Thoughts and Minds of Humankind. (A web article retrieved from ROSS through <https://rossintelligence.com/history-artificial-intelligence/> on September 19 2018).
- Colagrossi, M. 2018. What's the Difference between A. I., machine learning, and robotics? (Retrieved from <https://bigthink.com/mike-colagrossi/whats-the-difference-between-ai-machine-learning-and-robotics> on September 10 2018)
- Copeland, M. 2016. <https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/> retrieved on September 10 2018.
- Ganguly, D. S. 2018. Top Differences between Artificial Intelligence, Machine Learning & Deep Learning. (<https://hackernoon.com/top-differences-between-artificial-intelligence-machine-learning-deep-learning-d39cb6f6feaa> retrieved on September 10 2018).

- Geoghegan, B. D. 2008. The Historiographic Conceptualization of Information: A Critical Survey. Northwestern University. IEEE Annals of the History of Computing. Published by the IEEE Computer Society. (Retrieved from <https://pdfs.semanticscholar.org/9f5c/fd48578c0aceb817bfc8cc8ce65cfefe5de.pdf> on September 19 2018).
- Gomez-Herrera, E. 2013. Comparing Alternative Methods to Estimate Gravity Models of Bilateral Trade, *Empirical Economics*, Springer, Vol. 44(3), pages 1087-1111, June.
- Dalakov, G. 2018. History of Computers, Hardware, Software, Internet... (Accessed through <http://history-computer.com/index.html> on September 18 2018).
- Educalc.net. 2018. Blaise Pascal (1623-1662). (Web article) (Accessed from <https://www.educalc.net/196488.page> on September 18 2018).
- Einav, L. and J. D. Levin. 2013. The Data Revolution and Economic Analysis. National Bureau of Economic Research Working Paper 19035.
- Emant Pte Ltd. 2018. Logic Gates, Boolean Algebra, and Truth Tables. (Retrieved from <http://electronics-course.com/logic-gates> on September 18 2018).
- Erickson, J. 2014. Models of Computation. Lecture 6: Turing Machines. (A lecture note retrieved from <http://jeffe.cs.illinois.edu/teaching/algorithms/notes/models/06-turing-machines.pdf> on September 19 2018).
- Frede, D. and B. Reis (Eds.). 2009. *Body and Soul in Ancient Philosophy*. ISBN 978-3-11-020236-6.
- Klemen, J. 2007. From Artificial Neural Networks to Emotion Machines with Marvin Minsky. *Acta Polytechnical Hungarica*. Vol. 4 (4). (Retrieved from https://www.uni-obuda.hu/journal/Kelemen_12.pdf on September 20 2018).
- Lipton, Z. C. 2015. Deep Learning and the Triumph of Empiricism. (Retrieved from <https://www.kdnuggets.com/2015/07/deep-learning-triumph-empiricism-over-theoretical-mathematical-guarantees.html> in December 2018).
- Littman, M. L. 2006. Lecture 19: Neural Networks. CS105: Great Insights in Computer Science.
- Maliki, et. al. 2014. Neural Network Applications in the Forecasting of GDP of Nigeria as a Function of Key Stock Market Indicators.
- Morajda, J. 2003. Neural Networks and Their Economic Applications. In *Artificial Intelligence and Security in Computing Systems*, pp. 53–54, Springer, 2003.
- Negnevitsky, M. 2005. *Artificial Intelligence: A Guide to Intelligent Systems*. 2nd Ed. © Pearson Education Limited 2002. ISBN 0-321-20466-2.

- Oppy and Dowe. 2008. The Turing Test. Oppy and Dowe Encyclopedia of Philosophy. (Retrieved from <https://OppyandDowe.library.sydney.edu.au/archives/sum2010/entries/turing-test/> on September 17 2018).
- Piccinini, G. 2004. The First Computational Theory of Mind and Brain: A Close Look at McCulloch and Pitts's "Logical Calculus of Ideas Immanent in Nervous Activity". *Synthese*, 141(2), 175-215. (Retrieved from http://www.umsl.edu/~piccininig/First_Computational_Theory_of_Mind_and_Brain.pdf on 19 September 2018).
- Piketty, T. and E. Saez (2003). Income Inequality in the United States, 1913-1998. In *The Quarterly Journal of Economics*. Vol CXVIII Issue 1. February 2003. As cited by Einav and Levin 2013.
- Press, G. 2016. A Very Short History of Artificial Intelligence (AI). (Retrieved through <https://www.forbes.com/sites/gilpress/2016/12/30/a-very-short-history-of-artificial-intelligence-ai/#544eb8db6fba> in December 2018) .
- Skirry, J. n.d. Rene Descartes (1596-1650). A web article in: *Internet Encyclopedia of Philosophy*. (Accessed from <https://www.iep.utm.edu/descarte/> on September 17 2018).
- Smith, C., B. McGuire, T. Huang, and G. Yang. 2006. The History of Artificial Intelligence. University of Washington. (Retrieved from <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf> on September 19 2018).
- Statoo Consulting. 2016. What is Data Mining (Data Science)? (Retrieved from <http://www.statoo.com/en/datamining/> on 12 September 2018).
- Tillema, F., K. M. van Zuilekom, and M. F. A. M. van Maarseveen. 2006. Comparison of Neural Networks and Gravity Models in Trip Distribution. University of Twente, Department of Civil Engineering, Centre for Transport Studies, P.O. Box 217, 7500 AE, Enschede, The Netherlands.
- Titcomb. J. 2015. Who is George Boole and why is He Important? Today's Google Doodle Explained. In *The Telegraph*. (A web article retrieved through <https://www.telegraph.co.uk/technology/google/google-doodle/11968997/george-boole-google-doodle.html> on September 19 2018).
- Varian, H. R. 2014. Big Data: New Tricks for Econometrics.
- Verdu, S. 1998. Fifty Years of Shannon Theory. *IEEE Transactions on Information Theory*, Vol. 44, No. 6, October 1998. (Retrieved from <https://www.princeton.edu/~verdu/reprints/IT44.6.2057-2078.pdf> on September 19 2018).
- Villa, A. and V. Valaitis. 2018. Machine Learning Projection Methods for Macro-Finance Models.

- Wang and Raj. 2017. On the Origin of Deep Learning. Language Technologies Insitute, School of Computer Science, Carnegies Mellon University (Retrieved through <https://arxiv.org/pdf/1702.07800.pdf> in September 2018).
- Woelwar, A., J. P. Burgard, J. Kunst, and M. Vargas. 2018. Estimation Methods for Gravity Models (Package ‘gravity’).
- Wohl, I. and Kenedy, J. 2018. Neural Network Analysis of International Trade. USITC Office of Industries Working Paper ID-049. May 2018.
- Yeung, R. W. 2001. A First Course in Information Theory. The Chinese University of Hong Kong. Kluwer Academic Publishers. (A Draft).
- Yotov, Y.V. and R. Piermartini. 2016. Estimating Trade Policy Effects with Structural Gravity. WTO Working Paper ERSD-2016-10, 19 July 2016.