THE UNIVERSITY OF CHICAGO

The Relationship between Economic Freedom and Income Equality

By

Anqi Zhou

June 2021

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

Faculty Advisor: Philip Waggoner

Preceptor: Sanja Miklin

Abstract

This paper empirically examines the relationship between economic freedom and income inequality by using a machine-learning constructed economic freedom indicator that addresses three major methodological problems in the currently popular economic freedom indices. This paper selects the optimal Support Vector Regression (SVR) model with Radial Basis Function (RBF) kernel to construct the new economic freedom indicator, after comparing four different kernel types. The empirical results show that there exists a robust invert-U shaped relationship between economic freedom and income inequality.

keywords: economic freedom indicator, income inequality, machine learning, Support Vector Regression

1. Introduction

A large collection of literature on both the theory and the empirical evidence on the macroeconomic determinants of economic growth (including their effects on income inequality) has developed over the recent decades. At the same time, the effect of economic freedom on income inequality has continuously been a controversial topic in the empirical analysis. Both Berggren (1999) and Scully (2002) shows evidence that the relationship between economic freedom and income equality is complementary, not competitive. However, Carter (2007) reported there is a potential tradeoff between economic freedom and income equality. Therefore, the true underlying relationship between economic freedom and income inequality is a problem worth investigating, and its answer will present a meaningful guidance for government policy making.

There is at least one more reason why this question deserves more study: today's advanced techniques such as machine learning methods, are able to significantly improve the performances of economic freedom measurements, and thus will provide us with new insights and understandings regarding this question. Since concept of economic freedom is highly qualitative, the methods used to quantify it plays an important role in investigating its effects. In the past, Berggren (1999) and Scully (2002) used the Economic Freedom of the World (EFW) index, which was available for approximately 100 countries every five years from 1975 through 1990 (Gwartney, Lawson, & Block, 1996). Carter (2007) also chose a chain-linked version of the EFW index, which was available for 53 nations in 1970,

70 in 1975, 102 in 1980, 109 in 1985, 113 in 1990, 123 in 1995, and 123 in each year 2000 through 2003. However, commonly used economic freedom indices, including the EFW index, may not be a good choice to pursue for accuracy, as their methodology has been criticized by scholars since they were first established.

Therefore, the main contributions of this paper are twofold: 1) to construct a new, machine-learning-based, economic freedom indicator that addresses three major issues found in the commonly used indices' methodologies, in order to 2) revisit the problem of the relationship between economic freedom and income inequality. Based on this new economic freedom indicator, we establish a fixed effect regression model to analyze a panel data of 88 countries ranging from year 2000 to year 2018. The empirical results show a robust inverted-U relationship between economic freedom and income inequality, which suggests that economic freedom and income equality are competitive at a low range of economic freedom level but complementary at a high range of economic freedom level.

The rest of the paper is structured as follows. Section 2 discusses the general steps of constructing an economic freedom indicator. Section 3 introduces the current popular economic freedom indices and also considers the problems in their methodologies. Section 4 elaborates the construction of the new economic freedom indicator based on machine learning methods. I explain how to build the labelled dataset and the methods of feature extraction. I also discuss the choice of Support Vector Regression (SVR) model and compare optimal SVR models with different kernels. Lastly, I present the overview of new economic

freedom indicator. Section 5 examines the relationship between economic freedom level and income inequality. This section also includes the explanation of chosen variables and data sources. Section 6 concludes with closing remarks.

2. The Measurement of Economic Freedom

The basic approach to form a measurement for economic freedom level consists of three steps. The first step is to define the concept of economic freedom; the second step is to select a set of observable regime characteristics (or features), reflecting the key components of economic freedom concept; the third step is to transform these features into a measurement of economic freedom level.

1.1. Concept of Economic Freedom

In a frequently cited article, Gwartney, Lawson and Block (1996) provided an initial definition of economic freedom: "Individuals have economic freedom when (a) property they acquire without the use of force, fraud, or theft is protected from physical invasions by others, and (b) they are free to use, exchange, or give their property to another as long as their actions do not violate the identical rights of others." It mainly focuses on protection of individuals' property and the freedom of transactions.

But Gwartney, Lawson and Block (1996) also claim it is vital to distinguish political and civil rights from economic freedom, because they think there exists some circumstances in

which a country with a great protection of civil rights and a high level of democracy may still implement policies that severely limit economic freedom. However, much of the literature holds a different view. For instance, Wright (1982) mentions it is reasonable to accept that economic freedom and political and civil rights generally occur simultaneously, even though there may exist some uncommon cases. Furthermore, it is also reasonable to believe that the existence of political freedom and civil rights strengthens economic freedom, and this is particularly true in the area of legal structure.

Although scholars have not yet reached a consensus on a single definition of economic freedom, it seems that they have a broad consensus on what economic freedom is embodied as. (Keseljevic, 2013; Hanke and Walters, 1997; Gwartney, Block and Lawson 1996; Johnson, Holmes and Kirkpatrick, 1998) Generally speaking, scholars consider economic freedom to be embodied in the rule of law, property rights, political and civil rights, and freedom of contract, and to be characterized by external and internal openness of the market, the protection of rights and freedom of economic initiative.

1.2. Features Selection

The difficulty for this step is that most key components of economic freedom are not directly measurable, such as, trade freedom, market openness, property rights, etc. To address this problem, most of the literature used both the subjective indices and the objective regime characteristics to reflect each of the chosen dimensions of economic

freedom. For instance, in the components of Economic Freedom of the World (EFW) Index, "judicial independence" and "legal enforcement of contracts" are both the subjective indexes while "government consumption" is the objective regime characteristics. Those subjective indices generally include two types. One type is derived from the raw data of surveys which are designed by the experts, such as the "Judicial independence" index mentioned above. The "Judicial independence" index is from the *Global Competitiveness Report* question: "Is the judiciary in your country independent from political influences of members of government, citizens, or firms? No—heavily influenced (= 1) or Yes—entirely independent (= 7)". (The Fraser Institute, 2020, p.3) Another type is the one estimated by the experts using relevant raw data. For example, the "Legal enforcement of contracts" index is based on the World Bank's *Doing Business* estimates for the time and money required to collect a debt. The experts make assumptions on debt ratio and estimate the time and money cost value to gain this index. (The Fraser Institute, 2020) The former type depends on the design of survey questions and respondent's personal experience, while the latter type depends on the coders' personal experience or interests. Thus, both of them have a certain degree of subjectivity, but they generally provide a very complete dataset, for various countries and various years. In contrast, objective information of the regimes does not often exist for all regimes and period. Therefore, currently subjective indices, rather than objective regime characteristics, still dominate the literature.

1.3. Transformation

In general, every measurement of economic freedom needs to specify a function that maps the selected set of features into the level of economic freedom (EF). The function will be generalized as following:

$$EF_{i,t} = f(\boldsymbol{z}_{i,t}) \qquad \boldsymbol{z}_{i,t} = (z_{i,t}^1, z_{i,t}^2, \dots, z_{i,t}^n)$$

Where $\boldsymbol{z}_{i,t}$ is a vector of features for country *i* at the time point *t*, and *n* denotes the dimension of the vector.

The difficulty in this step, specifying the function, is that we cannot directly observe the true form of this function nor the true economic freedom level. The common way to address this problem, is to make assumptions on the function form. In the current popular economic freedom indices, which will be discussed later, this function is just assumed as the average function. However, the assumed function forms are often too simplistic and also lack theoretical support. So, one of the main contributions of this paper is that we use machine learning model to approximate this function without making prior assumptions on the form. But we need to first collect a dataset in which all the observations are correctly labelled. Then we try to use this dataset to reveal the true underlying relations between the regimes' relevant characteristics and their economic freedom level through machine learning methods. Specifically, we select Support Vector Machine Regression (SVR) model to capture the potential highly non-linear relations. The key advantages of SVR and the reasons we choose SVR for this case, will be discussed later in section 4.

3. Current Popular Indices for Economic Freedom

Today, there are two commonly used indices that systematically qualify economic freedom: The *Economic Freedom of the World* (EFW) indicator by Fraser Institute and *the Index of Economic Freedom* (IEF) by Heritage Foundation. Literature in the economic freedom topic usually applies either of them or make adjustments based on them to build the custom index that serve specific purposes. Although both indices take into account the same key elements of economic freedom (for example, security of property and freedom of exchange), they do have different emphases and methods.

3.1. Economic Freedom of the World (Fraser Institute)

In 1986, Economic Freedom of the World report from Fraser Institute designed an index that measures the degree to which the nations' institutions and policies are supportive of economic freedom. Today, the EFW indicator considers five fundamental areas: (1) "Size of Government", (2) "Legal System and Security of Property Rights"; (3) "Sound Money"; (4) "Freedom to Trade Internationally", and (5) "Regulation". This index measures economic freedom in five broad components which encompass 26 sub-components in total and thus include 44 variables. Each sub-component is a feature measured from 0 (completely unfreedom) to 10 (completely freedom).

We briefly introduce each component here to pave the way for the feature selection later. The sub-components constitution of the EFW index provides a good guideline for our

variable selection. For a complete list of the areas, sub-components, and underlying variables of EFW Index, please see Table A1 in Appendix A.

The first area, (1) "Size of Government", reflects the reliance on the Government and the political processes to allocate resources, goods, and services. (1A) "Government consumption" as a share of total consumption and (1B) "Transfers and subsidies" as a share of GDP indicates how much government spending is relative to spending of individuals, households and business units. (1C) "Government enterprises and investment" and (1E) "State ownership of assets" measure the proportion of government investment relative to private investment and enterprises. The reason for including (1D) "Top marginal tax rate" is that the high marginal tax rate applicable to relatively low-income levels also indicates dependence on the government. These five sub-components in area 1 are all calculated on empirical data.

The second area, (2) "Legal System and Property Rights" measures how effective the protection of persons and their acquired property is, which is one of the foundations of economic freedom. From (2A) "Judicial independence" to (2I) "Business costs of crime", the nine sub-components reflect various aspects of protective function and they are all subjective sub-indices coming from three primary sources, which are *Global Competitiveness Report*, *International Country Risk Guide*, and the World Bank's *Doing Business*. It's worth mentioning that (2J) "Gender Disparity Adjustment" measures the degree to which women and men obtain the same legal rights. It's used to make adjustments in area 2.

The third area, (3) “Sound Money” measures the degree of monetary stability and individual’s access to “sound” money. High and volatile inflation rates will hinder exchange by distorting the relative prices and erode the value of property. Therefore, sound money is also vital to economic freedom. (3A) “Money growth”, (3B) “Standard deviation of inflation”, and (3C) “Inflation: most recent year” are objective, measuring the consistency of monetary policy and long-term price stability. And (3D) “Freedom to own foreign currency bank accounts” is a subjective sub-index, measuring the ease of using other currencies through domestic and foreign bank accounts.

The fourth area, (4) “Freedom to Trade Internationally”, reflects various restraints that hinder international exchange, including the tariffs, the hidden regulatory restraints, the control on exchange rates and capital movement. While (4A) “Tariffs” are calculated using objective variables, (4B) “Regulatory trade barriers”, (4C) “Black-market exchange rates”, and (4D) “Controls of the movement of capital and people” are all constituted by subjective indices.

The fifth area, (5) “Regulation” measures the regulatory restraints that limit the freedom of exchange in credit, labor and product markets, thus this area consists of three sub-components: (5A) “Credit market regulations”, (5B) “Labor market regulations”, and (5C) “Business regulations”. The countries with less regulatory restraints towards these three markets will be rated higher. For each component in this area, there are several subjective sub-indices to describe various aspects.

3.2. Index of Economic Freedom (Heritage Foundation)

Another widely used index is The Index of Economic Freedom. It is created in 1995 by the Heritage Foundation and the Wall Street Journal, to measure the degree of freedom individuals have in controlling their own labor and property. This index is composed of four broad components: (1) "rule of law", (2) "government size", (3) "regulatory efficiency", and (4) "open markets". And in each board component, there are three sub-components. So, this index is measured by 12 quantitative and qualitative sub-components in total. Each sub-component is a feature measured from 0 to 100. Please see Table A2 in Appendix A for more information on the components and sub-components of IEF.

The Heritage Foundation website does not describe how these 12 sub-components are derived technically in detail, but it does describe the underlying representation of these 12 sub-components.

"Property rights" reflects how effective the protection towards property is and judicial effectiveness measures how effective the judicial system is. "Government integrity" shows indirectly the extent of corruption in government, for The Heritage Foundation believes corruption will hinder economic freedom.

The "Government size" component is very similar to the "Size of Government" in the EFW index, and it similarly reflects government spending and tax burden. "Fiscal health" measures the government's deficits and debt burden. This indicator is included because The

Heritage Foundation considers the budget as one of the clearest indicators of the extent to which the government respects the principle of limited government.

"Business freedom" measures the freedom in entrepreneurial activities. "Labor freedom" measures the freedom in the labor market, which basically describes the same aspects as those in the labor regulation sub-component of EFW index. "Monetary freedom" is characterized by a steady and reliable currency, and prices that are determined by markets (rather than by government control).

"Trade freedom" shows the freedom to trade internationally, similar to the respective concept in EFW index. "Investment freedom" requires the transparency and equity of supporting all types of companies, the encouragement of innovation and competition, the freedom of cross-border investment and capital movement. "Financial freedom" reflects how efficient and accessible the financial system is. The inefficiency of financial markets will increase the cost of financing activities and limit competition, which further hinders economic freedom.

3.3. The Discussion of Methodology Used by Current Popular Indices

The most common criticism regarding current popular economic freedom indices has to do with their methodology. Roughly speaking, there are three major problems in the methodology of these two popular indices: (1) factor weights (the form of transformation function), (2) composition of variables, (3) multicollinearity of components.

First, both the EFW indicator and the IEF average their components (features) with equal weights. However, this transformation function (taking the average) is most likely too simplistic to capture the true underlying relations between a set of features and economic freedom level. The true underlying relation is probably non-linear, not to mention the linearity that assign features with equal weights. The Heritage Foundation claims that IEF weights factors equally in order for the overall score to not be biased toward any one component or policy direction. (The Heritage Foundation, 2021) However, the scores of IEF are not normalized, so the index inherently weighs the factors because of the large variation in means.

Second, much of the literature, such as Heckelman and Stroup (2000, 2005), have mentioned that the composition of variables in the EFW index is too arbitrary. Rode (2012) used cluster analysis to regroup the 42 variables of EFW index into alternative categories and composed a quite different form of economic freedom index that reduce the multilinearity and reach the new conclusions on which part of economic freedom are responsible for economic growth. One of the alternative categories in Rode (2012), for example, is named "Government weight in the economy". It is regrouped by (1A) "Government consumption" sub-component, (1B) "Transfers and subsidies" sub-component, two of the variables underlying the (1D) "Top marginal tax rate" sub-component, three of the variables underlying the (5B) "Labor market regulations" sub-component and a sub-component named "size of the trade sector relative to expected"

(which does not exist in EFW index now). This may indicate that there is still much room for improvements of EFW Index's variable composition.

Third, both indices suffered from a severe multicollinearity problem. Dialga and Vallée (2018) stress that four components of IEF (property rights, business freedom, investment freedom and financial freedom) are strongly and positively correlated with each other, which leads to its lack of statistical validity. This kind of multicollinearity problem also exists in the EFW index. Rode (2012) reported the correlation coefficients between area 2, area 3, area 4 and area 5 of EFW index are all above 0.4, especially the correlation coefficient of area 2 and area 5 is more than 0.7.

Because of these potential problems, many scholars have criticized their indices and have been concerned that their application may lead to the insufficient or even wrong policy conclusions. (Keseljevic, 2013) Some scholars have also been making attempts to address the above issues. For instance, Dialga and Vallée (2018) used principal components analysis (PCA) and benefit of the doubt (BOD) methods to generate component-specific and country-specific weights, in order to recalculate the scores for IEF. Keseljevic (2013) used an Instrumental Variable-Two Stage Least Squares (IV-2SLS) empirical specification to report a new cross-country economic freedom ranking of 135 countries. This IV-2SLS specification is able to remove the original assumption of IEF that each component has equivalent effects on the overall indicator. However, few scholars have considered machine learning methods.

Therefore, the application of machine learning methods in this paper could provide us with a new perspective on how to construct the economic freedom indicator.

4. New Machine-Learning Constructed Index

4.1. Improvements

This paper constructs a new economic freedom indicator that addresses the three major problems of current popular indices. First, instead of using the mean as the underlying transformation function, this paper applies Support Vector Regression (SVR) model to identify a function which characterizes the underlying highly non-linear relations much better and thus provides a highly accurate outcome for economic freedom level. Second, this paper uses Principal Component Analysis (PCA) to perform dimension reduction on the set of regime characteristics variables. This method computes new principal components (PC) to capture the data's variation by transforming to a high dimension data space. These PCs are the variance-maximizing, orthogonal summaries of the data for our original variables. So, it naturally solves the arbitrary composition of variables problem. And we need to decide how many PCs we are going to use as features so that they are able to effectively capture sufficient variance in the data space. Features composed in this way are generally orthogonal and thus will not suffer from multicollinearity.

4.2. Methodology

4.2.1. Labelled Dataset

4.2.1.1. Labelled Dataset Construction

To train and test machine learning models, we need a dataset in which every data record (a country at a time point) has a set of feature values (regime characteristics) and a label value (its economic freedom level). It means regimes at particular time points can be included in this labelled dataset only if we can directly observe their economic freedom level without prior knowledge of the mapping function. To satisfy this requirement, as the starting point we select regimes with uncontroversial economic freedom levels. In this paper, we believe that the classification of countries whose economic freedom levels lie at the two extremes of the economic freedom spectrum is uncontroversial, as this means there is a consensus on the economic freedom levels of freest and least free countries. For example, Australia always ranks in the top decile of the rankings of current economic freedom indices, and it is indeed widely recognized by the public as highly economic free.

More specifically, we will classify a country-year observation as most economic free if the country ranks at top 15 in either EFW index or IEF in that year, while we will classify a country-year observation as least economic free if the country ranks at bottom 15 in either EFW index or IEF in that year. For observations of most economic freedom, we label their economic freedom level as 1 (the maximum), and for observations of least economic freedom, we label their economic freedom level as 0 (the minimum). In this way, we build an appropriate labelled dataset to start with.

4.2.1.2. Can the labelled dataset be regarded as consensus?

In order to construct a reliable indicator, the model requires us to provide a dataset that is correctly labelled as highly economic free or highly economic unfree. Therefore, whether there is a consensus in our labelled dataset is an important question. Next, we will explain the validity of our labelled dataset from various aspects.

First, all the countries in the labelled dataset have economic freedom levels that are consistent with our common sense (for a complete list of the country-year observations that have been selected into the labelled dataset, see Table A3 and Table A4 in Appendix B). There is no observation which has an unreasonable label and very few are debatable. For example, Australia, Austria, Canada, Chile, Denmark, Finland, Iceland, Ireland, New Zealand, Singapore, Sweden, and Switzerland are all considered as highly economic free regimes by the public. Algeria, Angola, Bolivia, Cameroon, Ethiopia, Gabon, Suriname, and Zimbabwe are always considered as highly economic unfree regimes by the public. Although Argentina was once one of the world's wealthiest countries and always impresses us as energetic and free (probably because of its famous football team), it is not the case in the economic freedom field. Argentina ranks the bottom in all of the three relevant indices, including EFW index, IEF and Global Competitive Index (which will be introduced later). The import and currency controls, the expropriation of firms in key sectors, the common corruption scandals, the complicated procedure of staring a business, the rigidity of its labor law all make it become one of the least economic free countries in the world.

Second, we test how other relevant indices assess the observations in our labelled datasets. Although there exist very few other indices measuring economic freedom, we did

find two relevant ones. The first one is the Scully-Slottje1991 index, which is developed in Scully and Slottje (1991) to measure economic freedom for countries (rather than country-year observations). The second one is the Global Competitive Index (GCI). It measures the "competitiveness" of a large number of countries' economies and the competitiveness is defined as "the set of institutions, policies, and factors that determine the level of productivity of a country." (Schwab, 2010, p.4) This index considers very similar aspects as economic freedom concept does, and it is categorized as a measure for economic liberty in Hanke and Walters (1997). Table 1 presents the consistency of labelled observation towards economic freedom level in the Scully-Slottje1991 index and the Global Competitive Index. Because I am only able to collect information about Global Competitive Index from 2007 to 2019, I only consider the labelled observations after 2007 when comparing with the Global Competitive Index. From Table 1, we will find there is a low degree of disagreement on our labelled dataset.

Table 1 - Consistency of Labelled Observations among Indices

	Most Economic Free Countries	
	Scully-Slottje1991 index (<=40)	Global Competitive Index (rank <=40)*
Overlap Rate (Coverage Rate)	80% (80%)	82.65% (100%)
	Least Economic Free Countries	
	Scully-Slottje1991 index (>=80)	Global Competitive Index (rank >= 80)*
Overlap Rate (Coverage Rate)	76.92% (81.25%)	95.83% (75%)

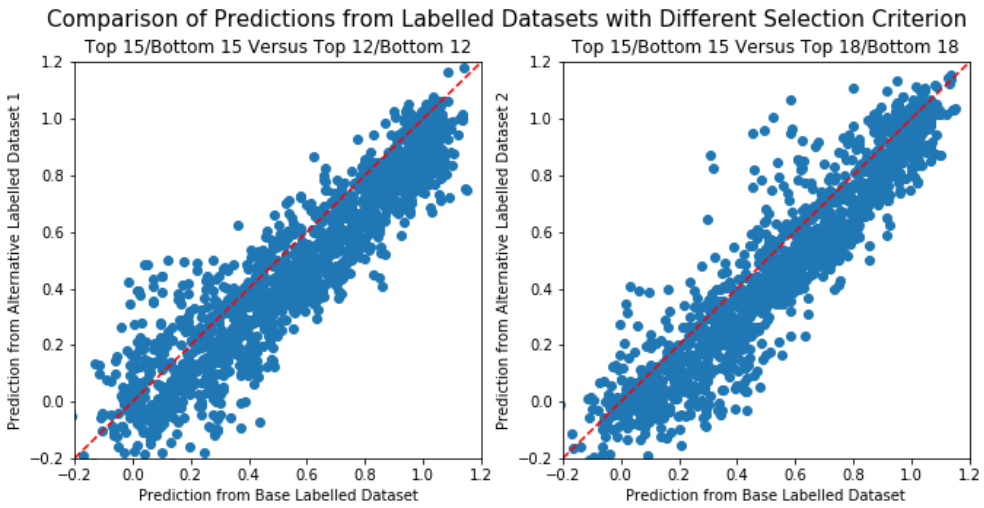
*only consider labelled observations after 2007

The two points discussed above provide evidence for consensus at the extremes.

4.2.1.3. Alternative Labelled Dataset

But how “extreme” should we choose? In fact, both the EFW index and IEF will also classify the most freedom and least freedom countries each year. However, their boundaries are very large, for instance, countries in the first quartile will all be regarded as most economic free in EFW index. Here I prefer to only select top 15 and bottom 15 countries each year which are most convincing, making sure I provide a dataset that are correctly labeled. Admitting the threshold I choose (top 15/bottom 15) is to some extent arbitrary, I thus test whether the index changes significantly when the labeling threshold changes. In the two graphs of Figure 1, each x-axis represents the index predicted based on the optimal SVR model trained by our baseline labelled dataset, which applies the top 15/bottom 15 labeling threshold. And in the first graph, the y-axis represents the index predicted from the optimal model trained by alternative labelled dataset 1 which applies the top 12/bottom 12 labeling threshold; while in the second graph, the y-axis represents the index predicted from the optimal model trained by alternative labelled dataset 2 which applies the top 18/bottom 18 labeling threshold. So, from Figure 1, we can directly see the comparison between the indexes that are constructed from three labelled datasets with different labelling thresholds.

Figure 1 – Comparison of Predictions from Labelled Dataset with Different Labelling Threshold



In both graphs of Figure 1, data points have been gathered around the 45-degree line, which suggests our index's notable robustness to changes in the labelling threshold. This robustness should not be surprising due to the consensus on the countries' economic freedom levels at the two extremes of the economic freedom spectrum.

4.2.2. Regime Characteristics and Feature Extraction

In previous section 1, we introduce the embodiment of economic freedom in a widely agreed manner. In summary, economic freedom is reflected in the rule of law, the protection of property rights, the protection of political and civil rights, the external and internal openness of the market, and the freedom of economic initiative. This has reported all the aspects that are widely accepted to indicate economic freedom level.

To some extent, the EFW index and IEF has told us how to measure these aspects of economic freedom. As we described in section 2, many components among EFW index and IEF are similar or overlap in meaning, and the measurements are analogous as well. To illustrate, Table 2 presents the respective aspect that each component or sub-component of

EFW and IEF describes, and the differences between the components of EFW and IEF that describe the same aspect.

Table 2 – Differences between Components of EFW and IEF

	EFW	IEF	Differences
Rule of Law Protection of Property Rights	Legal System and Property Rights	Rule of Law	EFW – measures legal system effectiveness more comprehensively. IEF – measures government integrity (corruption degree) in addition
External and Internal Openness of Market	Freedom to trade internationally	Trade Freedom	
Freedom of Economic Initiative	Regulation (Credit, Business, Labor)	Business Freedom Labor Freedom Investment Freedom Financial Freedom	EFW – measures freedom in credit market in addition IEF – measures freedom to investment and efficiency of financial system in addition
Government Intervene	Size of Government	Government Size	IEF – measures fiscal scale in addition
Monetary Stability	Sound Money	Monetary Freedom	IEF – measures price control in addition

Therefore, the natural idea is to incorporate all the above regime characteristics to the whole dataset initially and then construct proper features based on them for later prediction. Since these regime characteristics overlap in meaning, they suffer from a significant multicollinearity problem, which is evident in their VIF factors (see Appendix C). To solve this problem, the common way is to apply a feature extraction method to them for dimension reduction.

The feature extraction method this paper chooses is Principal Component Analysis (PCA). PCA is a widely used dimension reduction technique for making predictive model. It can increase interpretability and at the same time minimize information loss. The underlying idea is to linearly transform a set of potentially correlated variables' observations by orthogonal transformation, thereby projecting a set of uncorrelated variables' values. These transformed and uncorrelated variables are called principal

components (PC). Generally, we select the first few PCs that preserve most of the data's variation to be features for later prediction model.

Here, as we can see from Table 3, the first nine PCs have much more explanatory power in the data's variation (have much higher feature importance score) than other PCs. Therefore, we select them as the features for our later machine learning model. The regime characteristics underlying the nine PCs mainly include property rights, government integrity, government spending, investment freedom, tax burden, monetary freedom, trade freedom, financial freedom, investment freedom, business freedom, etc. It's surprising that most of these regime characteristics are variables from IEF. From this perspective, the variables constructed by IEF could reflect the data's variation well, although they still have a serious multicollinearity problem.

Table 3 – Feature Importance and Components

PC	PC Importance	Most Important	Second Important	Third Important
PC1	0.5358	Property_Rights, 0.491	Government_Integrity, 0.481	Government_Spending, 0.372
PC2	0.1478	Government_Spending, 0.825	Investment_Freedom, 0.307	Tax_Burden, 0.295
PC3	0.0663	Investment_Freedom, 0.654	Monetary_Freedom, 0.421	Government_Integrity, 0.411
PC4	0.0615	Trade_Freedom, 0.68	Financial_Freedom, 0.374	Property_Rights, 0.365
PC5	0.0449	Monetary_Freedom, 0.615	Tax_Burden, 0.565	Investment_Freedom, 0.374
PC6	0.0290	Trade_Freedom, 0.549	Monetary_Freedom, 0.507	Business_Freedom, 0.446
PC7	0.0280	Financial_Freedom, 0.678	Tax_Burden, 0.474	Investment_Freedom, 0.411
PC8	0.0259	Business_Freedom, 0.742	Tax_Burden, 0.338	Financial_Freedom, 0.33
PC9	0.0157	Government_Integrity, 0.671	Property_Rights, 0.661	Trade_Freedom, 0.213
PC10	0.0062	Own_Foreign_Accounts, 0.626	Conscription, 0.502	Capital_Controls, 0.342

4.2.3. Model - Support Vector Machine

This paper uses Support Vector Machine (SVM) algorithm for economic freedom level prediction. SVM is a supervised machine learning technique for pattern recognition, which

can be used to reveal the unknown relationship between a set of features and outcomes. (Steinwart and Christmann, 2008) And it is especially useful when the unknown relationship is complex. (Bennett and Campbell, 2000; Guenther and Schonlau, 2016) Unlike the conventional statistics methods (such as Logistic Regression), we do not need to specify the function in SVM that depicts the unknown relationship. So, one of SVM's most important features and advantages is that it is able to capture potential highly nonlinear relations through classification.

Another thing worth mentioning is that the SVM algorithm can be used for both classification and regression. Here we apply the SVM regression (also called Support Vector Regression) model, in order to predict a continuous economic freedom index. The Support Vector Regression uses the same principles as the SVM for classification, with only a few subtle differences. For example, SVR introduces the margin of tolerance (epsilon) to solve the infinite possibilities problem that comes from the fact of output being a real number. However, the underlying idea is basically the same: in order to minimize the error, find a hyperplane in the n-dimensional space that maximizes the margin (with a part of error tolerated).

In addition, there are four kernel function choices in the SVR model: linear, Radial Basis Function (RBF), polynomial and sigmoid. The kernel function generally transforms the low-dimensional data into high-dimensional data so that the data can be separated linearly in the high dimension space. That means we want to select the kernel type suitable for our

dataset (able to perform the linear separation on them after transformed by the selected kernel functions). To some extent, kernel type actually determines the shape form of hyperplanes.

The last thing is that the SVR model is very suitable for our case in this paper. SVR has some recognized advantages and disadvantages. On one hand, it is very generalizable, accurate, and it is robust to outliers. On the other hand, it is not suitable for large dataset, and does not work well when the number of features exceeds the number of training data samples. However, such characteristics makes SVR an appropriate choice for economic freedom level prediction in this paper. First of all, our dataset has a proper size to apply SVR model (with only 423 labelled data records and 9 features). Second, the generalizability and prediction accuracy are what we care most about the model, since our concern is to predict an economic freedom indicator for later investigation. For more reasons why we prefer the SVR model to the LR model in this paper, see Appendix D.

4.3. Results

4.3.1. MSE Comparison

This paper compares the best SVR models with different kernels, trying to find out the most appropriate kernel type. Here the best model refers to the model with optimal parameters, which is selected through a grid search with cross-validation. It turns out the one with RBF kernel has the least mean squared error (MSE) (see Appendix E). Therefore,

we are going to adopt the best SVR model with RBF kernel for the new economic freedom indicator construction.

4.3.2. Predictions for Test Dataset from Best Model

Here we present the prediction results of our best model (with RBF kernel) for test dataset, for a clear view at the model's accuracy. The labelled dataset is randomly split into the training and testing dataset, the ratios of which are 70% and 30% respectively. Thus, the training dataset contains 296 data records in total and the remaining 127 data records are used for testing. Figure 2 presents the boxplots of our best model's predictions towards economic freedom level for most economic free countries and least economic free countries in the test dataset, while Figure 3 shows the distributions of predictions towards economic freedom levels for the two groups of countries. Still for the same two groups, Figure 4 directly plots the prediction points in a single plot using different color for each of them.

Figure 2 – Boxplots of Economic Freedom Level Predictions for Most Economic Free and Least Economic Free Countries (Test Dataset)

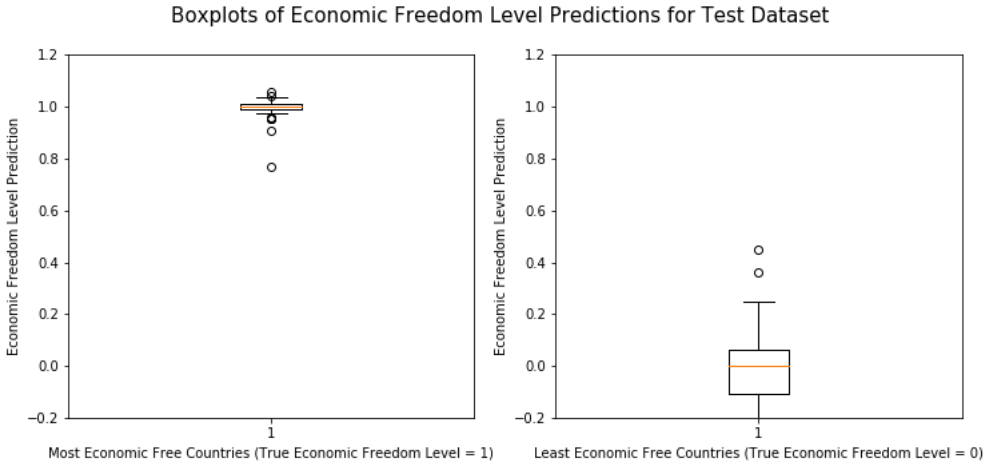


Figure 3 - Distributions of Economic Freedom Level Predictions for Most Economic Free and Least Economic Free Countries (Test Dataset)

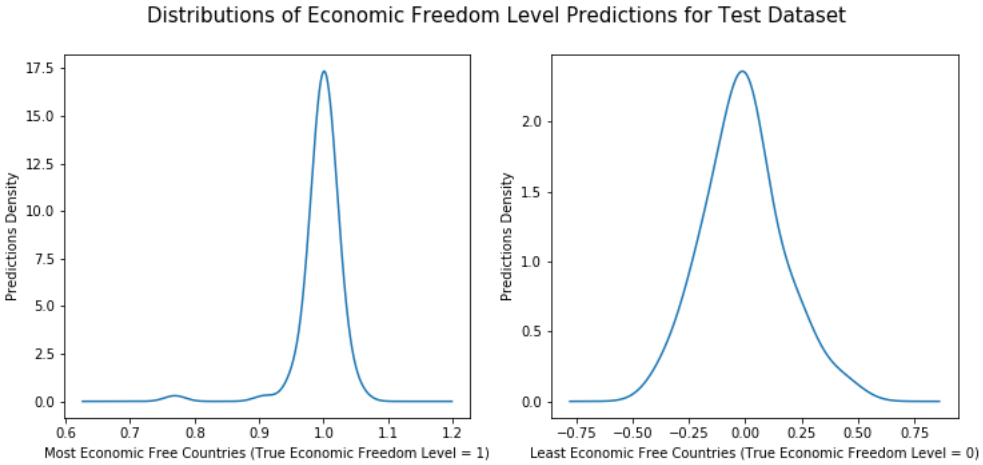
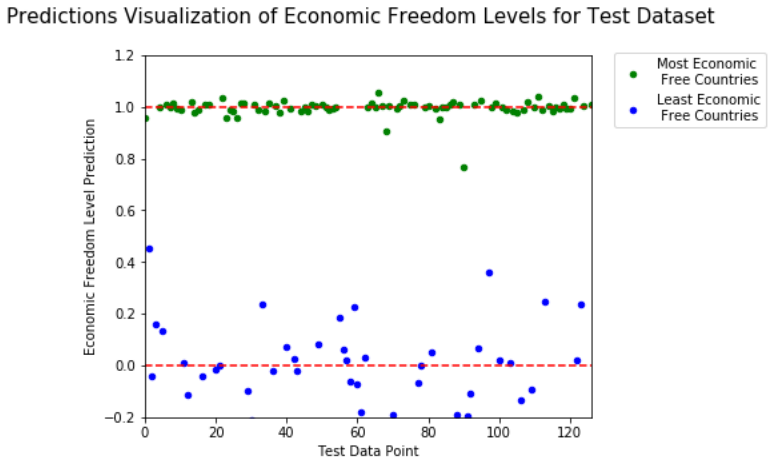


Figure 4 - Prediction Visualization of Economic Freedom Levels for Most Economic Free and Least Economic Free Countries (Test Dataset)



As we can see from these figures, the best model is highly accurate in predicting the most economic free countries, while it's not extremely accurate in predicting the least economic free countries. This is probably because the countries with complete economic freedom are always similarly free, but the countries with complete economic unfreedom have their own unfreedom aspects each. Just like the Chinese proverb, the happiness in happy families is always similar but the misfortunes vary differently in unfortunate families. But at least, the predictions of economic freedom level for least economic free countries still

falls in a very reasonable range – 75 percent quantile of predictions towards label = 0 is less than 0.0653. So, we can still conclude our best model has a high accuracy in prediction.

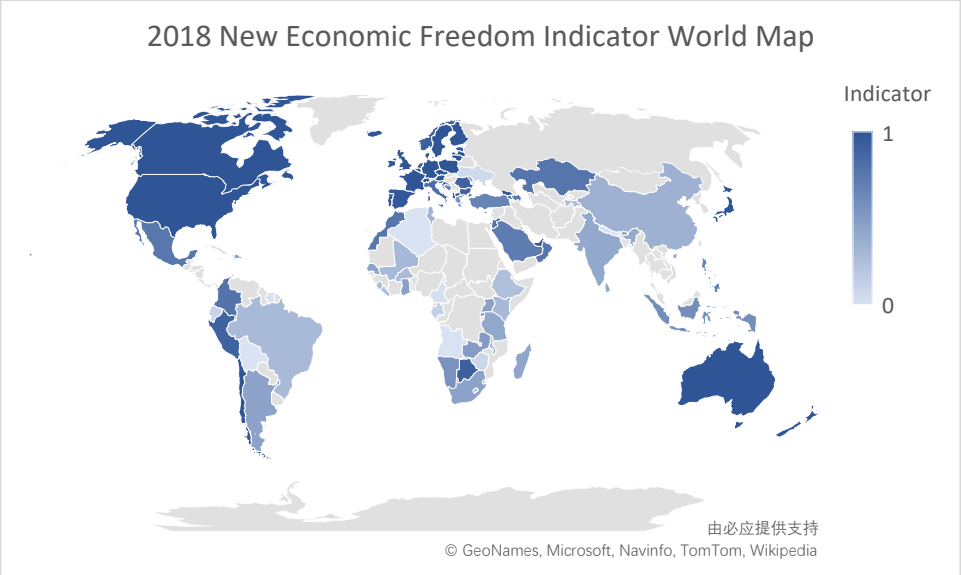
4.4. New Economic Freedom Indicator

4.4.1. Overview of New Economic Freedom Indicator

Finally, we apply our best model to the rest of the whole dataset to obtain the predictions of economic freedom levels for all the country-year observations from 2000 to 2018. We set the predictions higher than 1 as 1 and set the predictions lower than 0 as 0. In this way, we create a new economic freedom indicator based on machine learning methods. For more explanation on why we use this method to construct the new indicator instead of rescaling the predictions into 0-1 range, please see Appendix F.

Figure 5 presents an overview of economic freedom levels all around the world in year 2018. As we can see, most of North America, North Europe and Australia are of high economic freedom, while East Asia, most of Africa and South America have relatively low economic freedom. There seems to exist a geographical clustering for economic freedom.

Figure 5 – 2018 New Economic Freedom Indicator World Map



4.4.2. Distribution of New Economic Freedom Indicator

The first graph in Figure 6 plots the distributions of the new economic freedom indicator, EFW index and IEF in 2018. Because EFW Index ranges from 0 to 10 and IEF ranges from 0 to 100, here we divided EFW Index by 10 and IEF by 100 to plot the first graph. The values of EFW Index and IEF are very concentrated, partly because the actual ranges of the two indices are both smaller than the 0-1 range (for example, the lowest EFW Index value in 2018 is 0.475 but not 0). In contrast, our new economic freedom indicator set the minimum economic freedom level as 0 and set the maximum economic freedom level as 1. So, to eliminate the scaling effect, I rescale the EFW Index and IEF to the 0-1 range and plot the second graph. As we can see, the values of EFW Index and IEF are still much more concentrated than our new indicator. It's reasonable for us to doubt the similarity in the economic freedom levels of most countries around the world suggested by both EFW Index and IEF. This probably means our new indicator is able to capture the variation in countries' economic freedom levels better.

Figure 6 – Distributions of Economic Freedom Indices in Year 2018

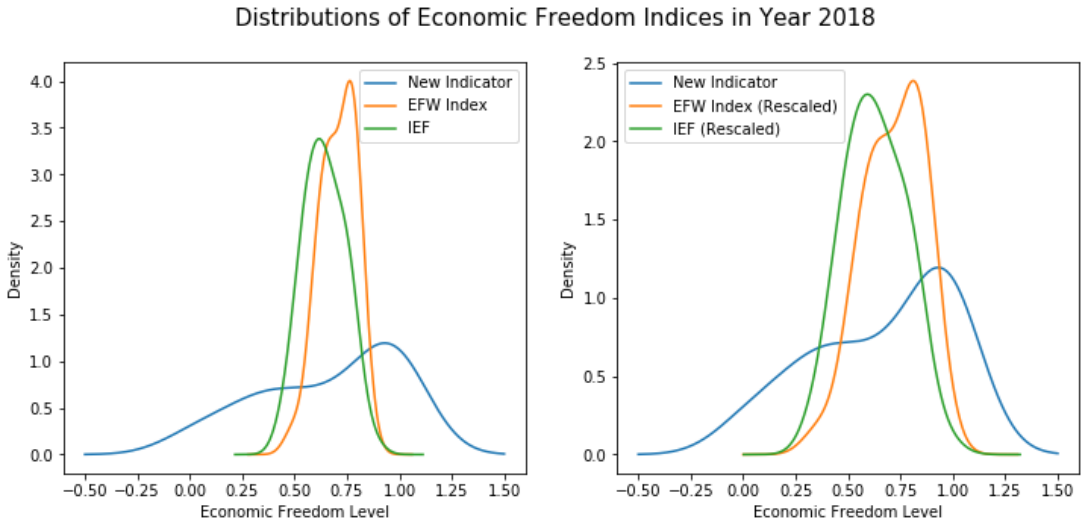


Figure 6 also addresses the concerns some readers may have on our model: will the prediction results be gathered in 0 or 1 as we use the binary-labelled dataset to train the model and it looks like we are teaching the model to distinguish 0 and 1? The answer is no. As we can see from Figure 6, there is no obvious polarization in the distribution of our indicator. The lack of popularization should not be surprising due to the powerful pattern recognition capabilities of the SVR model especially in the binary classification case. In theory, the distribution of our new indicator could possibly be close to the true economic freedom level distribution, since the SVR model has a very good generalization performance. But it's hard to prove this point empirically, as the true economic freedom level is unknown. All we can say is that the distribution is not contradictory to our common sense. There is a group of countries, probably developed countries in Europe and North America, that have sound systems and are highly economic free. There are other developing countries or countries in the Third World with varying levels of economic freedom, but

most of them lie in the middle of the economic freedom spectrum. Relatively few countries are extremely economic unfree.

5. Relationship between Economic Freedom and Income Inequality

5.3. Main Model Framework

The ultimate goal of this paper is to examine the relationship, which may not be linear, between economic freedom and income equity on a country level using new constructed economic freedom indicator. Kuznet's (1955) conjecture of an inverted-U relationship between inequality and per capita income has been supported by plenty of later literature. Therefore, I expect a similar non-linear Kuznets-like relationship might exist between economic freedom and income inequality, since the former has been widely shown to be positively related to economic growth. Bennett and Vedder (2013) also mention this possibility of Kuznets-like relationship and find relevant evidence using US-state level data. However, while more and more scholars recognize this relationship might be non linear, they never reached a consensus on its exact shape. Bennett and Vedder (2013) and Apergis and Cooray (2015) both suggest the existence of an inverted U-shaped relationship between economic freedom and income inequality, while Carter (2007) reach the opposite conclusion of a U-shaped relationship. Therefore, this paper will initially add quadratic terms of economic freedom to investigate the potential nonlinear relations. Hopefully, the new machine learning constructed economic freedom indicator may provide us with new understandings on this controversial issue. In addition, since Berggren (1999) suggested

the effects are different in the short run versus long run, lag terms of economic freedom will also be added trying to capture this difference.

Taking into account the lack of some country-level socio-economic data after 2000, this paper uses *ipolate* and *expolate* command in Stata to fill in the missing values by countries of all variables except for economic freedom indicator and Gini coefficient. The reason for not filling our key variables is to maximize our data's integrity. Then this paper establishes a fixed-effects regression model using the slightly unbalanced panel data of 88 countries from year 2000 to year 2018, in order to control the unobserved country-specific effects. Fixed-effects estimator is a widely used standard model to investigate macro-level laws in the field of economics. Much of the literature investigating the same topic also applies the FE estimator (Carter, 2007; Bergh and Nilsson, 2010; Sturm and De Haan, 2015). For more reasons why we prefer a fixed-effects model to a random-effects model, please see Appendix G.

In order to study this relationship between income inequality and economic freedom, we characterize the income inequality level by the Gini coefficient of each country as the dependent variable, as most literature does. Furthermore, the control variables mainly include aspects about income, education, demographics, workforce composition and urbanization degree, based on reference to the empirical models in Carter (2007) and Bennett and Nikolaev (2017). More specifically, the main model is as followings:

$$Gini_{i,t} = \beta_0 + \beta_1 Freedom_{i,t} + \beta_2 Freedom_{i,t}^2 + \beta_3 Freedom_{i,t-1} + \beta_4 M_{i,t} + \beta_5 Z_{i,t} + \varepsilon_{i,t}$$

where the subscript refers to country *i* at year *t*; $Gini_{i,t}$ is the Gini coefficient of country *i* at year *t*, representing the inequality level of a country; $Freedom_{i,t}$ is characterized by our new constructed economic freedom index; $M_{i,t}$ is a vector of standard economic explanatory variables, which have shown to be robustly linked with country-level inequality by previous empirical studies; $Z_{i,t}$ is a vector of possible additional economic explanatory variables, which may be related to country-level inequality based on previous literature; and $\varepsilon_{i,t}$ is the error term.

Although economic theory suggests a wide range of the potential determinants of income inequality, there is little consensus on the most relevant factors. (Furceri, Loungani and Ostry, 2019) Therefore, the variables in M vector only consist of mean years of schooling, but the variables in Z vector include per capita GDP, percent of population over age 15, percent of population over age 64, percent of employment in industry, percent of employment in services, percent of population in urban areas and fertilization rate. (Carter, 2007; Bennett and Nikolaev, 2017) Percent of population over age 15 (percent of youth population) and percent of population over age 64 (percent of elder population) depict a country's demographics; percent of employment in industry and percent of employment in services capture a country's composition of the workforce; percent of population in urban areas characterizes a country's degree of urbanization. Other potential determinants include trade openness, financial globalization and financial deregulation. (Furceri, Loungani and Ostry, 2019) But since they are actually involved in the economic freedom concept, they will not be put into the regression.

5.4. Data

5.4.1. Data Source for Gini Coefficients

This paper adopts the inequality measures (predicted Gini coefficients multiply by 100) from the latest version of Standardized World Income Inequality Database (SWIID), to characterize the inequality levels of country-year observations.

It employs data from Luxembourg Income Study as standard which provides high quality inequality measures, and also incorporates data from the *OECD Income Distribution Database*, the *Socio-Economic Database for Latin America & the Caribbean* generated by *CEDLAS* and the *World Bank*, *Eurostat*, the *World Bank's PovcalNet*, the *UN Economic Commission for Latin America & the Caribbean*, national statistical offices around the world and academic studies, which are basically the same data sources underlying the UNU/WIDER. But because SWIID's inequality measures are predicted based on the raw data, it is able to provide comparable measures for a large number of countries over a wide-ranging time. Furthermore, SWIID also provides the standard error for each measure, which is calculated by 100 imputations generated from Monte Carlo simulation. To ensure data's integrity, this paper only keeps the inequality measures with a standard error less than 3 (on a scale from 0 to 100).

5.4.2. Data Sources for Other Variables

The data for variables including per capita GDP, percent of population over age 15, percent of population over age 64, percent of employment in industry, percent of employment in services, percent of population in urban areas and fertilization rate, all come

from the *World Bank*, while the data for mean years of schooling variable comes from the *Our World in Data* dataset.

5.5. Result Analysis

5.5.1. Economic Freedom and Income Inequality

For robustness check, I ran eight fixed-effects regressions on the main model. Table 4 lists the entire estimation results.

Table 4 - Estimation Results

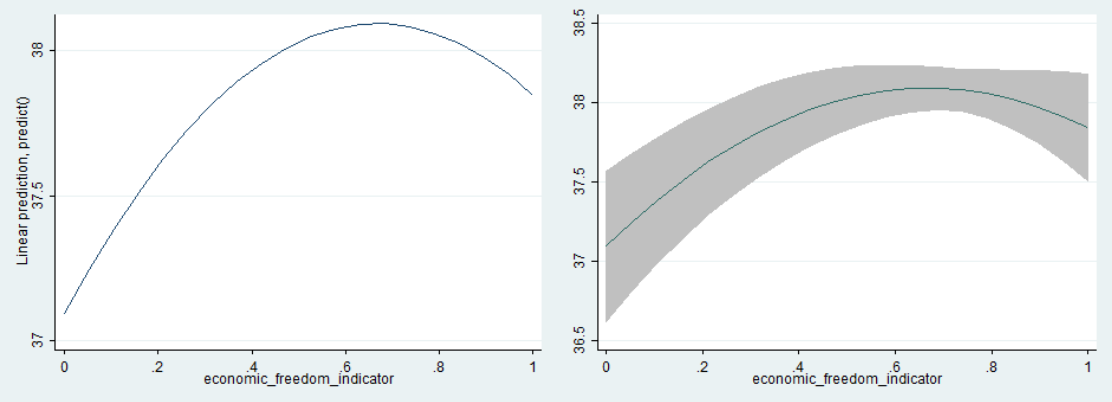
Dependent Variable: Gini coefficient								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
economic freedom	2.308***	1.953***	1.920***	1.726***	2.683***	2.961***	2.970***	2.987***
	(4.11)	(3.01)	(2.95)	(2.72)	(4.54)	(4.99)	(4.96)	(4.99)
economic freedom^2	-1.772***	-2.160***	-2.098***	-1.505***	-1.984***	-2.202***	-2.209***	-2.239***
	(-3.25)	(-3.87)	(-3.72)	(-2.71)	(-3.80)	(-4.22)	(-4.21)	(-4.27)
L.economic freedom		0.645	0.651	1.115***	1.180***	1.286***	1.289***	1.182***
		(1.54)	(1.55)	(2.70)	(3.11)	(3.39)	(3.39)	(3.09)
per capita GDP			-4.25e-06	2.61e-05***	1.65e-05***	1.38e-05**	1.38e-05**	1.1e-05
			(-0.75)	(4.07)	(2.75)	(2.31)	(2.30)	(1.79)
mean years of schooling				-0.710***	-0.382***	-0.447***	-0.447***	-0.409***
				(-9.21)	(-4.11)	(-4.47)	(-4.45)	(-4.01)
youth population percentage					0.411***	0.364***	0.363***	0.345***
					(14.10)	(12.01)	(11.6)	(10.56)
elder population percentage					0.327***	0.250***	0.251***	0.229***
					(7.58)	(5.33)	(5.26)	(4.67)
employment (%) in industry						-0.086***	-0.086***	-0.085***
						(-4.94)	(-4.88)	(-4.79)
employment (%) in services						-0.003	-0.003	3.15e-04
						(-0.25)	(-0.24)	(-0.02)
urban population percentage							-0.002	0.008
							(-0.11)	(0.41)
fertilization rate								0.350**
								(1.96)

** significant at 95% level.
*** significant at 99% level.

In all regressions, the economic freedom indicator and its quadratic term are both significant and consistent in their coefficients' signs and magnitude. This suggests that there exists a very robust inverted-U relationship between inequality and economic freedom, since the coefficient of squared economic freedom is negative. More specifically, income

inequality tends to increase at the initial stage of economic freedom and begins to fall as economic freedom continues to rise. Therefore, economic freedom and income equality are more like a trade-off relationship when the economic freedom level is relatively low, but these two can be achieved simultaneously after the economic freedom level becomes high. Figure 7 presents the marginal effects of economic freedom level (new economic freedom indicator) on income inequality (Gini coefficients) after controlling other variables (at their mean values). In Figure 7, the second graph has the same curve as the first graph does, but with a pointwise 95% confidence interval plotted in addition.

Figure 7 – Marginal Effects of Economic Freedom Indicator on Income Inequality (without or with a 95% Confidence Interval)



Then we could roughly calculate the turning point of this relationship based on regression coefficients. The coefficients of economic freedom level and its quadratic term robustly fall around 3 and 2.2 respectively. Thus, the turning point is about 0.67, which could also be clearly seen in Figure 7. The estimated marginal effects on income inequality changes from positive to negative at a new economic freedom indicator level of 0.67 (on a scale from 0 to 1). Income inequality continues to increase until economic freedom indicator

level exceeds a value around 0.67; once economic freedom indicator level increases to more than 0.67, the income inequality begins to decline. Consequently, countries with moderate economic freedom level (whose new economic freedom indicators are around 0.67) are more likely to exhibit the highest income inequality.

I also estimate the main model using the alternative new indicator (which is constructed by rescaling our SVR model predictions into the 0-1 range), the EFW index, and IEF. The estimation results from the models using alternative new indicator and IEF also suggest the existence of an inverted-U relationship, while the model using EFW index indicates the relationship is more like positively linear although this suggested relationship is not robust (see Appendix H for more details). So, the measurements of economic freedom could really have an impact on the analysis. The new economic freedom indicator, which I believe is a better alternative to the existing approaches, is able to provide us with new understandings on this question and also on the subsequent analysis.

5.5.2. Potential Mechanisms behind Inverted-U Relationship

There are two existing theories that could possibly explain the potential mechanisms behind this inverted-U relationship.

The first theory is discussed by Carter (2007). He believes economic freedom may have two kinds of effects on the income equality. On the one hand, economic freedom widens the opportunities of earning income for less privileged individuals, thus decreases the

inequality; on the other hand, economic freedom means less redistribution to the poor, via taxes, transfers and regulations, so that it may increase the inequality. Our regression outputs, which are contradictory to Carter (2007)'s results, indicate that in countries with relatively low economic freedom level, the second effect dominates the first effect when the economic freedom level increases; but in countries with relatively high economic freedom level, the first effect begins to exert greater power as economic freedom continues to rise.

The second theory is proposed by Bennett and Vedder (2013). Bennett and Vedder (2013) emphasizes the difference of economic opportunities at the low economic freedom level and the high economic freedom level. At the initial level of economic freedom, economic opportunities are very likely to come from launching an enterprise or engaging in trade and these opportunities will generally benefit the high-income population more than the low-income ones because only high-income people have the necessary physical or human capital to invest. As economic freedom level increases, the economy grows. However at this time, economic growth benefits the rich more, thus the income inequality is also raising. But when economic freedom continues to increase, new economic opportunities that can be used by people who lack capital, begin to emerge. And this eventually leads to a decrease in income inequality.

I believe the true mechanism behind the inverted-U relationship is very close to what Bennett and Vedder (2013)'s theory depicts, especially for the situation at a low level of economic freedom. In the countries with low economic freedom levels, the increase in

income inequality usually does not come from the reduction in redistribution to the poor through taxes, transfers and regulations, but mostly comes from the rich's monopoly on money-making opportunities. With the increase in economic freedom, the rich take the lead to seize the dividends of economic growth brought by the increase in economic freedom and leave little to the poor. But in countries with high economic freedom levels, the further increase in economic freedom benefits the low-income population because, thanks to the free market, the opportunities for making money and changing lives are now expanding even more.

5.5.3. Other Variables and Income Inequality

The estimation results from Table 4 also show us the relationships between other control variables and income inequality. After adding a mean years of schooling variable into the regression, the lag of economic freedom poses a strong and positive impact on income inequality. However, the impact is still less than concurrent economic freedom level. As for the income indicator, per capita GDP, both its impact level and significance decrease greatly after more control variables are taken into account. At the same time, the education indicator, mean years of schooling, has been proved to be robustly negative related to the income inequality, which suggests that high income equality is always associated with high education attainment. Positive and robust coefficients are shown as expected for both the percent of population under 15 and over 64, as this indicates higher income equality tends to be linked with higher percent of population in the prime of life. In addition, percent of

employment in industry also exerts a significantly negative impact on income inequality, while fertilization rate is positively related to inequality. Therefore, more workforce in industry and lower fertilization rate may be signs for higher income equality. Other variables including percent of employment in services and percent of population in urban areas do not seem to have robust impacts on income inequality.

6. Conclusion

A reliable indicator for economic freedom is vital for investigating and understanding the significance and effects of economic freedom. The machine learning methods to construct the indicator can basically solve the major limitations in methodologies of the current popular indices. Specifically, this paper uses the Support Vector Regression model with RBF kernel to develop a new reliable economic freedom indicator.

The relationship between income inequality and economic freedom are analyzed for a panel of about 88 countries from 2000 to 2018 based on the new indicator. The favorable effects on income equality include high education attainment, low percent of population under 15 and over 64, high percent of employment in industry and low fertilization rate. After these kinds of variables are held constant, the overall relationship between income inequality and economic freedom still exerts as a very robust inverted-U curve, which indicates that income equality and economic freedom are competitive at a low range of economic freedom level but complementary at a high range of economic freedom level.

7. Reference:

- Apergis, N., & Cooray, A. (2017). Economic freedom and income inequality: Evidence from a Panel of Global Economies—A linear and a non-linear long-run analysis. *The Manchester School, 85(*1), 88-105.
- Bennett, D. L., & Nikolaev, B. (2017). On the ambiguous economic freedom–inequality relationship. *Empirical Economics, 53*(2), 717-754.
- Bennett, D. L., & Vedder, R. K. (2013). A dynamic analysis of economic freedom and income inequality in the 50 US states: Empirical evidence of a parabolic relationship. *Journal of Regional Analysis & Policy, 43*(1), 42-55.
- Bennett, K. P., & Campbell, C. (2000). Support vector machines: hype or hallelujah?. *Acm Sigkdd Explorations Newsletter, 2*(2), 1-13.
- Berggren, N. (1999). Economic freedom and equality: Friends or foes?. *Public choice, 100*(3), 203-223.
- Bergh, A., & Nilsson, T. (2010). Do liberalization and globalization increase income inequality?. *European Journal of political economy*, 26(4), 488-505.
- Carter, J. R. (2007). An empirical note on economic freedom and income inequality. *Public Choice, 130*(1-2), 163-177.
- Dialga, I., & Vallée, T. (2018). The index of economic freedom: Methodological matters. *Studies in Economics and Finance*.
- Furceri, D., Loungani, P., & Ostry, J. D. (2019). The aggregate and distributional effects of financial globalization: Evidence from macro and sectoral data. *Journal of Money, Credit*

and Banking, 51, 163-198.

- Guenther, N., & Schonlau, M. (2016). Support vector machines. *The Stata Journal, 16*(4), 917-937.
- Gwartney, J. D., Lawson, R., & Block, W. (1996). *Economic freedom of the world*, 1975-1995. Fraser Inst.
- Hanke, S. H., & Walters, S. J. (1997). Economic freedom, prosperity, and equality: a survey. *Cato J.*, 17, 117.
- Heckelman, J. C., & Stroup, M. D. (2000). Whcich Economic Freedoms Contribute to Growth?. *Kyklos, 53*(4), 527-544.
- Heckelman, J. C., & Stroup, M. D. (2005). A comparison of aggregation methods for measures of economic freedom. *European Journal of Political Economy, 21*(4), 953-966.
- Johnson, B. T., Holmes, K. R., & Kirkpatrick, M. (1998). Heritage Foundation. *Wall Street Journal*.
- Keseljevic, A. (2013). Why Indexes of Economic Freedom Cannot Be Sufficient Indicators of Economic Freedom and Future Economic Prosperity?. *Available at SSRN 2238765*.
- Kuznets, S. (1955). Economic growth and income inequality. *The American economic review, 45*(1), 1-28.
- Rode, M. (2012). Do aggregation methods influence EFW Index performance? An examination using cluster analysis. *In University of Cantabria Working paper*.
- Scully, G. W., & Slottje, D. J. (1991). Ranking economic liberty across countries. *Public Choice, 69*(2), 121-152.

- Schwab, K. (2010, September). The global competitiveness report 2010-2011. *Geneva: World Economic Forum*.
- Scully, G. W. (2002). Economic freedom, government policy and the trade-off between equity and economic growth. *Public choice, 113*(1), 77-96.
- Steinwart, I., & Christmann, A. (2008). Support vector machines. *Springer Science & Business Media*.
- Sturm, J. E., & De Haan, J. (2015). Income inequality, capitalism, and ethno-linguistic fractionalization. *American Economic Review, 105*(5), 593-97.
- The Fraser Institute. (2020). *Composition of EFW Index*, 2020 [Table]. https://www.fraserinstitute.org/economic-freedom/approach
- The Fraser Institute. (2020). Explanatory Notes and Data Sources. *Economic Freedom of the World*. Retrieved from https://www.fraserinstitute.org/sites/default/files/economic-freedom-of-the-world-2019-appendix.pdf
- The Heritage Foundation. (2021). *Components and Sub-components of IEF*, 2021 [Table]. https://www.heritage.org/index/about
- The Heritage Foundation. (2021). About The Index. *2021 Index of Economic Freedom*. Retrieved from https://www.heritage.org/index/about
- Wright, L. M. (1982). A comparative survey of economic freedoms. *Freedom in the World: Political Rights and Civil Liberties*, 123-45.

Appendix A: Composition of EFW Index and IEF

Table A1 – Composition of EFW Index

1. Size of Government
A. Government consumption
B. Transfers and subsidies
C. Government enterprises and investment
D. Top marginal tax rate
(i) Top marginal income tax rate
(ii) Top marginal income and payroll tax rate
E. State ownership of assets
2. Legal System and Property Rights
A. Judicial independence
B. Impartial courts
C. Protection of property rights
D. Military interference in rule of law and politics
E. Integrity of the legal system
F. Legal enforcement of contracts
G. Regulatory costs of the sale of real property
H. Reliability of police
I. Business costs of crime
J. Gender Disparity Adjustment
3. Sound Money
A. Money growth
B. Standard deviation of inflation
C. Inflation: most recent year
D. Freedom to own foreign currency bank accounts
4.Freedom to Trade Internationally
A. Tariffs
(i) Revenue from trade taxes (% of trade sector)
(ii) Mean tariff rate
(iii) Standard deviation of tariff rates
B. Regulatory trade barriers
(i) Non-tariff trade barriers
(ii) Compliance costs of importing and exporting
C. Black-market exchange rates
D. Controls of the movement of capital and people
(i) Foreign ownership / investment restrictions
(ii) Capital controls
(iii) Freedom of foreigners to visit
5. Regulation
A. Credit market regulations
(i) Ownership of banks
(ii) Private sector credit
(iii) Interest rate controls negative real interest rates

B. Labor market regulations

(i) Hiring regulations and minimum wage

(ii) Hiring and firing regulations

(iii) Centralized collective bargaining

(iv) Hours regulations

(v) Mandated cost of worker dismissal

(vi) Conscription

C. Business regulations

(i) Administrative requirements

(ii) Bureaucracy costs

(iii) Starting a business

(iv) Extra payments / bribes / favoritism

(v) Licensing restrictions

(vi) Cost of tax compliance

Note. Adapted from https://www.fraserinstitute.org/economic-freedom/approach. Copyright 2020 by The Fraser Institute.

Table A2 - Components and Sub-components of IEF

rule of law	**government size**	**regulatory efficiency**	**open markets**
property rights	government spending	business freedom	trade freedom
government integrity	tax burden	labor freedom	investment freedom
judicial effectiveness	**fiscal health**	**monetary freedom**	**financial freedom**

Note. Adapted from https://www.heritage.org/index/about. Copyright 2021 by The Heritage Foundation.

Appendix B: Country-year Observations in the Labelled Dataset

Table A3 – Country-year Observation with Label = 0 in Labelled Dataset

country	year
Algeria	2000-2018
Angola	2000-2018
Argentina	2012-2016
Bolivia	2017, 2018
Bosnia and Herzegovina	2002-2004
Cameroon	2002, 2007, 2008
Ethiopia	2005-2016
Gabon	2010-2013
Guinea-Bissau	2013
Malawi	2000-2006
Mali	2002, 2004
Rwanda	2000
Sierra Leone	2000-2007
Suriname	2017-2018
Ukraine	2000, 2001, 2015, 2017
Zimbabwe	2000-2018

Table A4 – Country-year Observation with Label = 1 in Labelled Dataset

country	year
Australia	2000-2018
Austria	2000-2005
Bahrain	2000-2014
Canada	2000-2018
Chile	2000-2018
Denmark	2000-2018
Estonia	2001-2018
Finland	2000-2012
Georgia	2006-2018
Germany	2002, 2013, 2015
Iceland	2000-2009, 2018
Ireland	2000-2018
Japan	2000-2008
Latvia	2006-2017
Lithuania	2014-2018
Mauritius	2009-2018
Netherlands	2000, 2002-2012
New Zealand	2000-2018
Peru	2010
Singapore	2000-2018
Spain	2003, 2004
Sweden	2010, 2018
Switzerland	2000-2018
United Arab Emirates	2000-2018
United States	2000-2018

Appendix C: VIF Factors of Regime Characteristics from EFW Index and IEF

Table A5 – VIF Factors of Regime Characteristics

Regime Characteristics	VIF Factor
Intercept	674.8756
Government Consumption	5.314332
Transfers and Subsidies	7.206541
Government Investment	3.245715
Top Marginal Income Tax Rate	9.917914
Top Marginal Income and Payroll Tax Rate	7.723085
State Ownership of Assets	11.25649
Judicial Independence	48.27611
Impartial Courts	55.79719
Protection of Property Rights	15.08019
Military Interference in Rule of Law and Politics	6.779569
Integrity of the Legal System	12.09124
Legal Enforcement of Contracts	10.80471
Regulatory Restrictions on the Sale of Real Property	6.314592
Reliability of Police	11.92719
Gender Legal Rights Adjustment	4.445
Money Growth	4.61512
Standard Deviation of Inflation	5.985544
Inflation Most Recent Year	4.874877
Own Foreign Accounts	11.91279
Revenue From Trade Taxes	3.770948
Mean Tariff Rate	12.75669
Standard Deviation of Tariff Rates	3.432182
Non-tariff Trade Barriers	5.977647
Compliance Costs of Importing and Exporting	7.144255
Black Market Exchange Rates	2.709346
Financial Openness	16.47194
Capital Controls	5.167215
Ownership of Banks	5.771349
Private Sector Credit	2.013663
Interest Rate Controls	6.878342
Hiring Regulations and Minimum Wage	2.638025
Hiring and Firing Regulations	4.101421
Centralized Collective Bargaining	3.66863
Hours Regulations	3.202434
Mandated Cost of Worker Dismissal	4.585475
Conscription	2.660447
Administrative Requirements	4.013723
Regulatory Burden	4.272366
Starting A Business	9.176525
Impartial Public Administration	12.97585

Tax Compliance	5.582438
Property Rights	31.05449
Government Integrity	36.10433
Tax Burden	8.233194
Government Spending	5.183088
Business Freedom	9.87913
Monetary Freedom	6.595847
Trade Freedom	11.55997
Investment Freedom	7.734961
Financial Freedom	8.319525

Appendix D: Justification for Choice of SVR Model over LR Model

Our training dataset is binary labelled. So, it's natural to think about applying Logistic Regression (LR) model. However, next we will explain SVR has many absolute advantages over LR in the case of economic freedom level prediction for this paper.

The difference in their loss functions is the essential difference between these two models, which in fact reflects the difference of their underlying classification principles. SVR uses hinge loss function, while LR uses log loss function. The purpose of both loss functions is to increase the weight of data points with greater impact and decrease the weight of data points with less relevance. However, SVR's loss function is to only consider support vectors, the few points most relevant to classification, while LR's is to greatly reduce the weight of points far away from the hyperplane through nonlinear mapping, so as to relatively increase the weight of the most relevant data points.

Therefore, to some extent, SVR only considers data points on the hyperplane, while LR considers all data points. That's why SVR is far more robust to outliers than LR. So, in theory, SVR will provide much more robust predictions for economic freedom level than LR, especially in the area around 0.5 - a small change would cause the predictions around 0.5 from LR model shift strongly towards 0 or 1.

Second, as we can see, hinge loss function has a flat zero piece which makes the solution of the support vector machine sparse, while log loss function is a smooth and monotonically decreasing function. This results in that LR's solution depends on more training data samples and thus LR's sample overhead for prediction is much higher.

Empirically speaking, LR model performs much better in large datasets than in small datasets. A relatively small dataset like our labelled dataset is more suitable to apply to an SVR model.

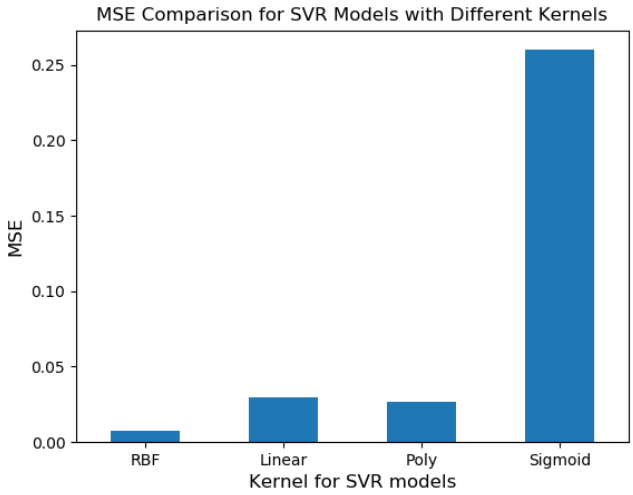
Third, SVR's strong generalization ability is exactly what we want for prediction purpose. As we can notice, SVR's optimization function already includes structural risk minimization, but LR's optimization is only to minimize empirical risk when the regularization term has not been added. Therefore, theoretically speaking, the SVR model has a stronger generalization ability than the LR model.

Appendix E: MSE Comparison Results

Table A6 – MSE Comparison for Best Models with Different Kernels

Kernel Type	MSE
RBF	0.00783
Linear	0.02962
Polynomial	0.02658
Sigmoid	0.25990

Figure A1 – MSE Comparison for Best SVR Models with Different Kernels



Appendix F: Why Assigning the Predictions Outside of the 0-1 Range to 0 or 1 Instead of Rescaling the Predictions?

In our model, the training data only consists of the regimes whose economic freedom level is maximum or minimum. Using such dataset, the prediction result derived from the model can represent the absolute level of a country's economic freedom, instead of just being an ordinal indicator (representing the ranking). Therefore, it is reasonable to regard the countries whose economic freedom predictions are larger than 1 as the most economic free countries and thus set their economic freedom index as 1. The same is for the countries with predictions less than 0. It is worth noting that I also considered rescaling the predictions to the 0-1 range, so as to create the new index. Since the predictions represent the absolute economic freedom levels, I believe both methods are acceptable. But in this paper, I prefer the former method which avoids violating my prior belief that the observations in the labelled dataset are the extremes of the economic freedom spectrum.

Furthermore, I also tried to use the alternative new indicator which is rescaled from the model predictions to estimate the main model. The results are consistent with our results in section 5 (see Appendix H for more details). Therefore, the choice between these two methods should not be a big problem.

Appendix G: Why Fixed Effects but not Random Effects for Main Model?

The fixed-effects (FE) model allows the heterogeneity intercept u_i to be correlated with the explanatory variables $X_{i,t}$; in contrast, random-effects (RE) model has a stricter assumption which is more difficult to satisfy, that is, u_i should not be correlated with any $X_{i,t}$. For this case, RE is definitely not appropriate, because the null hypothesis of the Hausman test that country-specific effects are not correlated with the explanatories variables has been easily rejected (chi-square = 47.07, p = 0.000). Since the Hausman test does not hold, the parameter estimates obtained by RE will be not consistent. But regardless of whether the Hausman test holds or not, the FE's estimates are always consistent. (It's just that the RE's estimates will be more efficient if the Hausman test holds.)

Another thing worth mentioning is that the researchers in the field of economics generally use the FE model, even though some statistical tests (such as the Hausman test) can tell us to a certain extent whether we should use FE or RE. Economists often pay attention to the macro-level laws of cities, states, or countries. Obviously, such a grouping variable (e.g., country) itself can represent its population (that is, the countries in the world are generally fixed). So, it's not a random sampling from the population. At the same time, a country's "baseline level" u_i are always related with other explanatory variables. Therefore, FE is more suitable for most cases in the economics field.

Appendix H: Estimation Results of Main Model Using Different Economic Freedom Indices

Table A7 reports the estimation results of the main model using the alternative new economic freedom indicator (which is constructed by rescaling our SVR model predictions into the 0-1 range), the EFW Index and IEF. Notice that here we divided the EFW Index by 10 and IEF by 100.

Table A7 - Estimation Results of Main Model Using Different Economic Freedom Indices

Dependent Variable: Gini coefficient			
	(1) (Alternative New Indicator)	(2) (EFW Index)	(3) (IEF)
economic freedom	3.998***	13.176	50.402***
	(3.72)	(1.12)	(6.31)
economic freedom^2	-2.553***	-6.031	-37.934***
	(-2.76)	(-0.70)	(-5.92)
L.economic freedom	1.585***	8.470***	8.639***
	(3.08)	(4.03)	(5.17)
per capita GDP	1.02e-05	1.08e-05	6.86e-06
	(1.66)	(1.78)	(1.15)
mean years of schooling	-0.406***	-0.317***	-0.318***
	(-3.96)	(-3.04)	(-3.23)
youth population percentage	0.346***	0.407***	0.353***
	(10.55)	(11.20)	(11.25)
elder population percentage	0.213***	0.175***	0.171***
	(4.33)	(3.54)	(3.72)
employment (%) in industry	-0.082***	-0.056***	-0.085***
	(-4.64)	(-2.94)	(-5.03)
employment (%) in services	0.002	0.029**	-0.008
	(0.14)	(2.00)	(-0.60)
urban population percentage	0.015	0.024	0.049***
	(0.79)	(1.23)	(2.68)
fertilization rate	0.427**	1.014***	0.367**
	(2.38)	(5.33)	(2.14)

** significant at 95% level.
*** significant at 99% level.

Figure A2 plots the marginal effects of economic freedom on income inequality in these three models after holding other control variables at their mean values. The x-axes of three graphs in Figure A2 are the alternative new indicator, the EFW index and IEF respectively, while the y-axes all represent the income inequality. The models using the alternative new

economic freedom indicator and IEF both suggest a robust inverted-U relationship between economic freedom and income inequality, although the curve's turning point and shape could be a little different from our results in section 5. At the same time, the model using the EFW index indicates that economic freedom and income equality is more like a competitive relationship, but the relationship is not robust as neither economic freedom nor its quadratic term is significant in this model.

Figure A2 – Marginal Effects of Economic Freedom on Income Inequality Shown in Three Models