MASTERS THESIS

The Impact of Trade, Comparative Advantage, and Sector Productivities on the Rate of Structural Transformation: An Empirical Analysis of the Three-Sector Ricardian Model

Alessandra Guerrero

ANR: 757064

Tilburg University

School of Economics and Management

Msc. Economics

Supervised by Dr. Guzman Ourens

Word Count: 11,845

20th of August, 2021

ABSTRACT

Introduction. Structural transformation is the process of the reallocation of resources across different sectors in the economy. Recently, policy makers have focused on structural transformation as a key aspect of economic growth, and potentially of economic deterioration.

Research Question. How is structural transformation (measures by employment in a sector) affected by measures of trade (the presence and extent of trade), measures of sectoral labor productivity, and measures of comparative advantage (both endogenous and exogenous measures)? To what extent is structural transformation caused by each of these factors?

Data. I use data from various databases: WIOD, GGDC-10, PWT, and World Bank Development Indicators. Using existing theoretical frameworks, I extract the necessary measures of economic factors that are key in structural transformation.

Methodology. Various econometric models are implemented. First is the Pooled OLS regression, followed by the Fixed-Effects (FE) and Random-Effects (RE) regressions. I construct a robust covariance matrix estimator for the chosen RE specification, and implement that RE specification onto the group of high-income countries represented in the sample.

Results. The impact of bilateral trade shares, shares of value added in gross output, and geographical and demographic factors are both large in magnitude and highly statistically significance. The income group implementation demonstrates that different countries demonstrate the same pattern of results. Trade, sector specific productivity, and geographical variables are key determinants of the changes in the shares of total labor in each sector.

Conclusions. Theoretical considerations of the importance of trade, comparative advantage, and sector productivity are verified in this research. Structural changes in employment by sector are sensitive to small percentage changes in each of these factors, and policy makers must be aware of that sensitivity when shifting those factors.

Key words: *structural transformation, macroeconometrics, fixed-effect models, Ricardian trade model, comparative advantage, econometric model, unobserved heterogeneity, time series econometrics.*

1. INTRODUCTION

Structural transformation is the process of reallocation of economic activity across different sectors of the economy. In other words, the reallocation of employment and productivity from one sector to another. Economists typically aggregate these numerous sectors into three primary sectors: agriculture, industry (manufacturing), and service. Structural transformation - otherwise known as structural change - has been consistently cited as one of the key aspects of economic growth, and as such has been the recent subject of potential policy debate. Developments in the study of structural transformation have shifted toward Ricardian trade models and trade openness considerations as additions to the vast literature of general equilibrium models, and thus have brought into question the importance of both international trade and comparative advantage on structural change.

Structural change - the reallocation of economic activity across the sectors of agriculture, industry, and services as we define it here - is increasingly one of the most important features of economic growth. Dudley Seers (1963) first brought to the field the concept of studying industrial characteristics of countries; specifically, how to associate differences in industrial characteristics (and characteristics across sectors) with differences in the level of development of a country. Following Seers, Kuznets (1973) presented structural transformation as one of the six main features of modern economic growth, and essential in the study of economic growth in less developed countries. In 2011, McMillan and Rodrik (2011) verified the claims of both Seers and Kuznets, and demonstrated that the reallocation of labor (employment) and other inputs to more productive sectors or subsectors was vital to economic development. They note, in addition, that structural transformation can also be equally detrimental to the economy as it can be beneficial from the misallocation of inputs to non-productive activities.

It is clear then that structural transformation serves as an important factor to consider in the creation and implementation of policy interventions, particularly for countries which are less developed or on the course of industrialization and development. Other economists who focus on the study of structural transformation, such as Herrendorf, Roberson, and Valentinyi (2013) concur with the conclusions of their predecessors, and furthermore note that the concept of the reallocation of economic activity at the sectoral level has received significant attention within policy circles and has “led to calls” for policy interventions (both in more developed and less developed countries).

Analysis of shifting resources between sectors, both empirically and theoretically, has revealed the three key facts of structural transformation: the linear decline in the agricultural sector's share across levels of development, the quadratic hump shape of the industrial sector's share, and the linear rise in the service sector's share as a country’s income increases.

More recent studies, such as those from Teignier (2018) and Swiecki (2017), stress the importance of international trade and trade openness in bolstering and an accelerating structural transformation. The combination of these key facts and the role of trade - particularly the trends of deindustrialization (the decline of manufacturing), continuous decrease in agriculture in the last couple of decades, and rise in low-skilled labor in services - lead to concerns over possible policy responses to potentially negative economic outcomes, as emphasized by van Neuss (2019). These studies are primarily reliant on a general equilibrium underpinning, and a basis in a three-sector Ricardian trade model. This model allows for the necessary inclusion of trade and the examination of comparative advantages in agriculture, industry, and services, and has been shown to be beneficial in the study of structural change. Hence the role of trade and productivity – in addition to comparative advantage measures – must be studied within an empirical framework to strengthen the results of its theoretical predecessors.

This research will seek to answer the question of whether trade, comparative advantage, and sector productivities – through several economic and country-specific measures – have an effect on structural transformation, represented as the share of total employment in agriculture, industry (manufacturing), and service sectors. Clearly, comparative advantage in a certain sector and the presence of trade would shift the allocation of labor from a less productive sector to this more productive sector. For illustrate this mechanism, the presence of extensive agricultural land could be a predictor of a country’s comparative advantage in the agricultural sector; if the country trades freely, and the country has a comparative advantage in agriculture, we can expect the share of trade in that sector to be similarly large. This would mean that it would be more efficient for the country to put more inputs (e.g. labor) into the agriculture sector, and hence we may see a rise in the share of employment in agriculture. This means that labor was reallocated

3. DATA & DESCRIPTIVE STATISTICS

3.1 THE SAMPLE & DATA SOURCES

To conduct this empirical analysis, I constructed a panel dataset from various databases. The World Input-Output Database (henceforth WIOD) provides data on bilateral trade, value added, gross output, and intermediate inputs by country and sector – as seen in Timmer et al. (2014) – for 39 countries and 34 sectors. The GGDC 10-Sector Database and the EU Klems Database provide data on value added in current prices, value added in constant prices, and total employment by sector. The GGDC Productivity Level Database (see Inklaar and Timmer (2008)) provides data on prices across the same sectors of the WIOD in 2005. From the countries in the WIOD, the Penn World Table from Feenstra et al. (2015) provides data on expenditure-side GDP and population data. Lastly, the World Bank Development Indicators database provides country-specific data on several geographical, crop, agriculture, and population variables. These serve as measures of comparative advantage and country-specific controls, and include population growth in rural and urban areas, agricultural land use, crop production, and GDP growth).

Using the model specifications, calculations, and derivations from Sposi (2019) and Swiecki (2017), I combine the data available in 10 subsectors into three primary sectors as is standard in structural change literature: agriculture, industry (manufacturing), and services. Following the calculations of both Sposi and Swiecki, I calculated the variables of interest in this research work (see the corresponding appendices of each paper for the specific theoretical derivations):

1) *Bilateral Trade Shares* – Sposi (2019) constructs bilateral trade shares for each country pair, in each of the three sectors, as the trade flows from an exporting country to an importing country divided by the importing country's gross output minus exports (within the same sector). This is the measure of trade used in this work, as it tells us the share of bilateral trade flows between two countries that belong to each sector. The unit is percent of total real bilateral trade flows between all three sectors.
2) *Share of Value Added in Gross Output* – Sposi (2019) computes the share of value added for a sector in a country as the value added in that sector divided by the gross output of that sector (within that country). Value added in gross output is the

difference between the gross output and the cost of intermediate inputs, and hence serves as a productivity measure of any entity to the economy (in this case, a specific sector). The unit is percent of total value added between all three sectors.

3) *Share of Total Employment* – I calculated the share of total employment for each sector by diving total employment in each sector by the total employment in each country for each year. Each of the shares of total employment in agriculture, industry, and services are the dependent variables in this study. The unit is percent of total employment.
4) *Urban Population Growth* – As per the World Bank, "Urban population refers to people living in urban areas as defined by national statistical offices." The World Bank explains that the growth of cities and urban areas signifies the shift from an agriculture-based economy to an industrial, or service-focused economy. Hence, it can be a crucial variable to include in the analysis of structural change, particularly as a measure of comparative advantage. The unit is in annual percent growth of the urban population.
5) *Rural Population Growth* – The World Bank defines the rural population as "people living in rural areas as defined by national statistical offices." It explains that where population densities are low – typically rural areas – markets are very thin, and the unit cost of infrastructure and all types of economic services are very high. This has large implications for development, and thus along with urban population growth, is an important measure of comparative advantage. The unit is in annual percent growth of the rural population.
6) *Total Land Area* – A country's total area, excluding bodies of water such as major rivers and lakes. The World Bank states that land area is particularly important for "understand an economy's agricultural capacity." Because of its importance to potential resources and potential productivity in the use of those natural resources, it is a key measure of comparative advantage in this study. The unit is square kilometers.
7) *Food Production Index* – The food production index covers food crops that are edible and contain nutrients, therefore excluding coffee and tea and inedible commodities. It is prepared by the Food and Agriculture Organization of the United Nations (FAO),

which calculate the indices of agricultural production in relative level of the aggregate volume of agricultural production for each year compared to that of 2014-2016 (=100) using a weighted average quantities of different food commodities.[1] Along with total land area and rural population growth, the food production index serves as a possible indicator of comparative advantage in agriculture. The unit is in index points.

8) *Crop Production Index* – Like the food production index, the crop production index shows agricultural production for each year relative to the base period 2014-2016 (=100). The commodities covered in this computation are all the crop and livestock products originating in each country, excluding food crops, seeds, and fodder. In the same vein as the above variables, this is as well an important possible indicator of comparative advantage, as it includes all crops (including oil and livestock) and not just food. The unit is also in index points.
9) *Agricultural Land* – According to the World Bank, agricultural land "refers to the share of land area that is arable, under permanent crops, and under permanent pastures." As such, this measure includes arable land. It does, however, exclude land under trees grown for wood or timber. It is important to include both total land area and agricultural land, and total land area includes areas not suitable for agriculture, and thus agricultural land dictates the potential availability of natural resources or crop production. It is crucial for understanding the movements of the agricultural sector, and hence a vital measure of comparative advantage. The unit is in percent of total land area.

The original balanced panel dataset was built using a sample of 38 countries included in the WIOD tables, excluding Taiwan and the Rest of the World (ISO code ROW) due to a consistent lack of data in the World Bank Development Indicator database. However, due to missing observations within the dependent variables (share of total employment in each of the three sectors), I restricted the sample by removing several countries from the dataset: Bulgaria (BGR), Canada (CAN), South Korea (KOR), Romania (ROM), Russia (RUS), and Turkey

[1] See https://databank.worldbank.org/metadataglossary/wdi-database-archives-(beta)/series/AG.PRD.FOOD.XD for more information on the calculation of the index.

(TUR). In addition, some years within this restricted dataset contained missing employment data, hence some countries do not have data for one or two years.

The cleaned panel dataset is composed of the variables discussed in this section, calculated from the above databases, for the following countries and each of the three sectors, and from 1995-2011 (the completeness of data within the time period depends on the country):

AUS	Australia	FIN	Finland	LTU	Lithuania	SWE	Sweden
AUT	Austria	FRA	France	LUX	Luxembourg	USA	United States
BEL	Belgium	GBR	United Kingdom	LVA	Latvia		
BRA	Brazil	GRC	Greece	MEX	Mexico		
CHN	China	HUN	Hungary	MLT	Malta		
CYP	Cyprus	IDN	Indonesia	NLD	Netherlands		
CZE	Czech Republic	IND	India	POL	Poland		
DEU	Germany	IRL	Ireland	PRT	Portugal		
DNK	Denmark	ITA	Italy	SVK	Slovakia		
ESP	Spain	JPN	Japan	SVN	Slovenia		

Table 1. List of Countries by ISO Code

It must be mentioned that this sample is not representative of the population of all countries in the world experiencing structural change. Out of the 32 countries represented in this sample, 2 are in the lower-middle income group, 3 are in the upper-middle income group, and the remaining 28 are in the high income groups. Only 5 of the countries are not member countries of the OECD, and 22 are European countries. No African countries are represented. This sample is not so representative due to the highly variable availability of data in lower income countries, or untrustworthy data presented by different parties (e.g. the central government) within these countries. All of the countries presented, however, have experienced structural transformation to some extent, considering the majority are high income, highly developed countries.

Since the original dataset contains multiples of country-year data points due to the bilateral trade share data presented in importing-exporting country pairs for every year (thus resulting in 25,194 observations), I took the weighted average of the bilateral trade shares of all the importing-exporting data points for each sector and year. Hence, I dealt with the issue of non-unique observations, and compacted the dataset from 25,194 observations to 663 observations.

4. EMPIRICAL ANALYSIS

4.1 ECONOMETRIC METHODOLOGY

Panel datasets are characterized as a collection of data from the same subjects (referred to as groups) over a period of time such that there exists a unique observation for each subject or group for each given period of time. In this study, I have an observation for each of the 33 countries at each year for the years 1995-2011. Hence, this data has both individual factors from the characteristics of a cross-sectional dataset and time factors from the characteristics of a time-series dataset, all contained within a panel dataset. .

The benefits of utilizing panel data lie primarily in its ability to allow a researcher to account for endogeneity caused by unobserved heterogeneity across subjects. Unobserved heterogeneity refers to the unobserved relationships between the independent variables within a regression model and any other independent variables not included in the model. This leads, without any robustness corrections, to a non-zero correlation between the error term of the regression (which represents all information not presented within the model) and the independent variables of the model. Not accounting for these within a study leads to biased estimation results and potential overestimation of a coefficient estimate. Using panel data allows the researcher to account for both of these issues. Thus, the panel dataset makes it feasible to more accurately analyze the effects of changes in various economic factors on structural transformation in this research by measuring the changes in the same countries over time.

There are three different regressions that are done on panel data. First is the Pooled OLS model, which ignores any of the time or individual characteristics of the observations and only measures the relationships between each subject. This is the basic panel data regression, represented by Equation 1 below:

$$Y_{it} = \beta_0 + \beta_1 X_{it,1} + \beta_2 X_{it,2} + \ldots + \beta_k X_{it,k} \ + \ \alpha_i + u_{it} \quad \text{……………………………..Eq. 1}$$

Where i represents each of the 39 countries (each subject) and $t = 1, 2, \ldots, 17$ represents each time period in years from 1995-2011. Y_{it} is the dependent (outcome) variable, $X_{it,1}, X_{it,2}, \ldots, X_{it,k}$ are the independent variables, $\beta_1, \beta_2, \ldots, \beta_k$ are the coefficients corresponding to the independent variables, α_i are the individual (i.e. fixed) effects, and u_{it} the idiosyncratic error term. The Pooled OLS model (as an OLS regression) requires the fulfillment of the conditions exogeneity and zero correlation between any unobserved independent variables and independent variables in the model. In addition, the Pooled OLS ignores the presence of serial correlation in α_i, and hence will give biased results and is inappropriate for panel data. I will provide the Pooled OLS estimation results as the first-line naïve regression results.

The second model for panel data is the Fixed-Effects (FE) Model, which will make it possible to determine the effects of the unobserved independent variables as a constant over each time period. Since the individual unobserved heterogeneity is equal across time, it can be ignored, and hence heterogeneity within individuals can be controlled. Here, α_i is assumed to be constant, and if one subtracts the mean values from each variable term (time-demeaning each term), then $\alpha_i = 0$ and can be ignored. The FE model is represented by Equation 2 below:

$$Y_{it} - \bar{Y_i} = \beta_1\left(X_{it,1} - \bar{X_i}\right) + \beta_2\left(X_{it,2} - \bar{X_i}\right) + \ldots + \beta_k\left(X_{it,k} - \bar{X_i}\right) + \left(u_{it} - \bar{u_i}\right) \text{ ....Eq. 2}$$

The idiosyncratic error must still be exogenous and non-collinear with any of the independent variables. The benefit of the FE model is that it permits the existence of heterogeneity within each subject (country in this case). Since the time-fixed (individual) effects of the data disappear, we can apply an OLS regression to derive the FE estimator for each variable.[2]

Lastly is the Random Effects (RE) model, which determines the time-fixed effects (individual effects) of the independent variables not within the model as random variables over the given time period. The RE model assumes that α_i is uncorrelated with any of the independent variables. If the covariance between α_i and the independent variables is zero, then there is no correlation and POLS is preferred. If it is not zero, then the FE model is estimated. The RE model then determines which of the two models to utilize according to the presence of serial correlation of the idiosyncratic error term, using a new term λ, which calculates how large the variance of α_i is. Equation 3 describes this λ term:

$$\lambda = 1 - \left(\frac{\sigma_u^2}{\sigma_u^2 + k \cdot \sigma_u^2}\right) \text{ ....Eq. 3}$$

If λ is 0 then the POLS is preferred. If λ is 1then the FE model is preferred. To determine which of the three models is more appropriate for this panel data using the F-Test and the

[2] See Toward Data Science for more detailed explanations not presented here.

Hausman Test. In addition, I will test for cross-sectional dependence, serial correlation, stationarity, and heteroskedasticity to reach the most appropriate model.

4.2 REGRESSIONS FOR EMPLOYMENT IN AGRICULTURE

Table 4 below presents the three separate estimation models for changes in employment in the agricultural sector. I estimate the Pooled OLS (POLS), Fixed Effects (FE), and Random Effects (RE) models, and include the coefficients and standard errors. *(Note: The standard errors are included below the coefficient estimates. The variables are abbreviated, and the constants excluded.)*

Variable	***POLS***	***FE***	***RE***
Bi. Trade Shares (A)	7.79*** (1.06)	2.07*** (0.51)	2.20*** (0.53)
Bi. Trade Shares (M)	- 3.62*** (0.99)	1.11*** (0.32)	1.30*** (0.32)
Bi. Trade Shares (S)	12.68* (5.55)	4.61** (1.65)	4.77*** (1.73)
S. Value Added (A)	2.27*** (0.23)	0.28** (0.11)	0.34** (0.11)
S. Value Added (M)	-1.55*** (0.24)	0.42*** (0.12)	0.42** (0.12)
S. Value Added (S)	0.42 (0.30)	0.11 (0.26)	-0.04*** (0.26)
Urban Pop. Growth	2.39*** (0.42)	-0.21 (0.19)	-0.21 (0.19)
Rural Pop. Growth	-0.30** (0.10)	-0.00 (0.03)	0.003 (0.03)
Total Land Area	0.08*** (0.019)	8.90 (19.39)	0.24*** (0.05)
GDP Growth	0.177 (0.106)	0.01 (0.03)	0.01 (0.04)
Food Prod. Index	0.63*** (0.45)	-0.69*** (0.15)	-0.65*** (0.16)
Crop Prod. Index	-2.30*** (0.34)	0.44*** (0.11)	0.42*** (0.12)
Agricultural Land	0.03 (0.05)	0.90*** (0.14)	0.61*** (0.12)
Number of Groups = 33 Number of Time Periods = 8-17 Number of Observations = 464 *** 1% significance level; ** 5% significance level; * 10% significance level; constant excluded.			

Table 4: Employment in Agriculture as the Outcome Variable, 1995-2011

As I will later discuss on page 22, the variables of interest are those within the Random Effects (RE) column. One can see that all trade and productivity variables, as well as the geographic measures of comparative advantage are highly statistically significant (at either the 1% or 5% significance level). Using the conclusions of previous studies on structural transformation as presented in Section 1, we can expect bilateral trade shares and shares of value added in agriculture to have a positive effect on the share of total employment in the agricultural sector. We can also expect the geographic factors to have a positive effect as well. Since I took the natural log of each variable, we can read the effect of each variable on the share of total employment in each sector as a percentage change.

Beginning with the measures of trade, as expected, a 1% increase in the bilateral trade share in agriculture results in a 2.20% increase in the share of total labor that goes into the agricultural sector. Similarly, increasing the bilateral trade shares (henceforth trade) in industry and services by 1% increases employment in agriculture by 1.30% and 4.77%, respectively. Looking at the shares of value added, we can see that an increase of the share of value added in agriculture of 1% leads to a 0.34% increase in employment in agriculture. That was also expected considering the proposed mechanism as explained in Section 2. Employment in agriculture increases by 0.42% when the share of value added in industry increases by 1%, and decreases by 0.04% when the share of value added in services increases by 1%. Although the positive coefficient on the share of value added (henceforth productivity) in industry may be counterintuitive in terms of the typical reallocation of labor from agriculture to industry, it could be a sign that the manufacturing sector is heavily reliant on agricultural production. Hence, if productivity improves in manufacturing it could require more labor in agriculture.

Likewise, the coefficient on productivity in services can make sense, as services and agriculture may not be closely linked as industry and agriculture. Hence, increasing productivity in services would decrease employment in agriculture. Generally, it appears that trade has large positive effects on employment in agriculture.

As for the measures of comparative advantage, and as we would expect, most measures of agricultural capacity are positively related to increases in agricultural labor. A 1% increase in total land area results in a 0.24% increase in agricultural labor. A 1% increase in the crop production index and in agricultural land use result in a 0.42% and 0.61% increase in labor in agriculture, respectively. On the other hand, a 1% increase in the food production index results in a 0.65% decrease in agricultural employment. Although this result is not expected, and I foresaw the food production index having a similar effect on the agricultural sector as the crop production index, it could be that an increased food production index has more of a positive effect on industrial employment due to increased food processing. That remains to be seen in the regression for the share of total employment in agriculture. In all, it appears that comparative advantage, only in the form of geographical variables, play a key role in the share of total employment in agriculture (and hence, structural transformation).

The next step is to find which of the three models is more suitable for this data. To do this, I complete an F-Test for individual effects to discern if the FE model is preferred to the POLS model. Given the nature of the panel dataset, we expect the FE model to be a better fit. The null hypothesis of the F-test is that the POLS model is preferred to the FE. Table 5 below presents the results of the F-Test:

Value	***Hypothesis***	***Result***
F-Statistic = 404.61 DF1 = 32 DF2 = 418 P-Value < 0.000	*Null*: POLS > FE *Alternative*: FE > POLS	*P-Value is significantly smaller than zero, we reject the null hypothesis.*

Table 5: F-Statistic Test for Employment in Agriculture

According to the results, the FE model is preferred to the POLS due to significant present individual effects, as was expected. The last step in choosing the correct model is to detect if the FE model is preferred to the RE model. I do this by utilizing the Hausman test. In the Hausman test, the null hypothesis is that the RE model is preferred to the FE model, and tests if the errors of the regression equations are correlated with the independent variables (in the null hypothesis they are not). Since RE models are estimated with partial pooling, then the results are more appropriate for few data points in a group, and the estimated coefficients come from a distribution from larger groups (that is, more countries in the world not included in the sample). Table 6 below presents the results of the Hausman test:

Value	***Hypothesis***	***Result***
Chi-Squ. = 6.15 DF = 13 P-Value = 0.940	*Null*: RE > FE *Alternative*: FE > RE	*P-Value is significantly larger than zero, we fail to reject the null hypothesis.*

Table 6: Hausman Test for Employment in Agriculture

According to these results, the RE model is preferred to the FE model for this data. Therefore, since the FE model is preferable to the OLS model, and the RE model is preferable to

Again, as with employment in agriculture and industry, the RE model is the most appropriate model. I need only consider the RE model for the next steps in the panel analysis for all three sectors.

4.5 DIAGNOSTIC TESTS AND ROBUST RESULTS

Now that the appropriate panel data model is chosen, I can perform a series of diagnostic tests to arrive at an optimal and robust econometric model for employment in each sector.[4] The theoretical underpinnings of these tests are not important to the results of this study, so they will not be discussed in within the scope of this paper.

I will begin with a test on cross-sectional dependence (contemporaneous correlation), which tests the correlation of residual errors across subjects, as correlation of these residuals could lead to biased estimation results. There are two possible tests to accomplish this goal. According to Baltagi (2005), the Breusch-Pagan LM test (the first test) does not perform well for panels with more observations than time periods; however, Pasaran's CD test performs equal well for a large number of observations and small number of time periods (as is in the case of this research, with 663 observations and 17 time periods). Therefore, Table 13 presents the results of Pasaran's CD test on the three FE regressions.

The null hypothesis is that there is no evidence of cross-sectional dependence. The results of Pasaran's CD test demonstrates that all three RE models of employment in agriculture, industry, and services show evidence of cross-sectional dependence.

Model	***Value***	***Hypothesis***	***Result***
RE: Employment in Agriculture	Z-Value = 8.75 P-Value < 000	*Null*: No cross-sectional dependence. *Alternative*: Cross-sectional dependence detected.	*P-Value significantly smaller than zero, reject null hypothesis.*
RE: Employment in Industry	Z-Value = 15.03 P-Value < 0.000	*Null*: No cross-sectional dependence. *Alternative*: Cross-sectional dependence detected.	*P-Value significantly smaller than 0, reject null hypothesis.*
RE: Employment in Services	Z-Value = 4.93 P-Value < 0.000	*Null*: No cross-sectional dependence. *Alternative*: Cross-sectional dependence detected.	*P-Value significantly smaller than zero, reject null hypothesis.*

Table 13: Results of Pasaran's CD Test on the Three RE Models.

[4] More detail on diagnostics on Fixed/Random Effects Models by Oscar Torres-Reyna.

In addition to testing cross-sectional dependence, it is important to test the three models for serial correlation. Serial correlation tests are common in macroeconomic panel datasets with moderately long time series, as is the case with the panel dataset for this research. The null hypothesis of the Breusch-Godfrey/Woolridge test is that there is no serial correlation, and one can expect with this type of data that there will be evidence of serial correlation. Table 14 presents the results of this test on each of the three RE models:

Model	***Value***	***Hypothesis***	***Result***
RE: Employment in Agriculture	Chi-Squ. = 194.37 DF = 8 P-Value < 0.000	*Null*: No serial correlation. *Alternative*: Serial correlation.	*P-Value significantly smaller than zero, reject null hypothesis.*
RE: Employment in Industry	Chi-Squ. = 193.02 DF = 8 P-Value < 0.000	*Null*: No serial correlation. *Alternative*: Serial correlation.	*P-Value significantly smaller than zero, reject null hypothesis.*
RE: Employment in Services	Chi-Squ. = 206.81 DF = 8 P-Value < 0.000	*Null*: No serial correlation. *Alternative*: Serial correlation.	*P-Value significantly smaller than zero, reject null hypothesis.*

Table 14: Results of Breusch-Godfrey/Woolridge Test on the Three RE Models.

As expected, the results show that there is evidence of serial correlation within all three RE models. This does not pose an econometric problem, as it can be readily addressed in implementing a robust model with the appropriate specifications to avoid biased estimates.

Last is testing for heteroskedasticity by utilizing the Breusch-Pagan test. Identifying if a model exhibits heteroskedasticity is important, as heteroskedasticity that is uncounted for leads to incorrectly estimated standard errors. The null hypothesis for the test is the presence of homoskedasticity. Both heteroskedasticity and serial correlation can be addressed by implementing a robust covariance matrix (robust standard errors) into the RE model, so the presence of heteroskedasticity is not cause for concern. Table 15 on the next page presents the results of the Breusch-Pagan test on the three RE models:

Model	***Value***	***Hypothesis***	***Result***
RE: Employment in Agriculture	BP = 425.67 DF = 45 P-Value < 0.000	*Null*: Homoskedasticity. *Alternative*: Heteroskedasticity.	*P-Value significantly smaller than zero, reject null hypothesis.*
RE: Employment in Industry	BP = 214.98 DF = 58 P-Value < 0.000	*Null*: Homoskedasticity. *Alternative*: Heteroskedasticity.	*P-Value significantly smaller than zero, reject null hypothesis.*
RE: Employment in Services	BP = 428.24 DF = 45 P-Value < 0.000	*Null*: Homoskedasticity. *Alternative*: Heteroskedasticity.	*P-Value significantly smaller than zero, reject null hypothesis.*

Table 15: Results of Breusch-Pagan Test on the Three RE Models.

The results of the test show that all three RE models on the share of total employment demonstrate the presence of heteroskedasticity. The diagnostic tests on the three RE models are completed, and I can now build a more appropriate model for employment in each sector through robust covariance matrix estimation. One can control for the presence of serial correlation and heteroskedasticity (as well as cross-sectional dependence) within a RE model through the use of a “Sandwich” estimator. Calculating the White standard errors based on the variance-covariance matrix allows the robust model to account for both heteroskedasticity and serial correlation within RE models, under the assumption of consistent heteroskedasticity, as seen in Kleiber and Zeileis (2008).

Table 16 on the next page presents these robust regression results:

Variable	***RE: Agriculture (A)***	***RE: Industry (M)***	***RE: Services (S)***
Bi. Trade Shares (A)	2.20*** (0.64)	0.20 (0.13)	-0.45* (0.22)
Bi. Trade Shares (M)	1.30** (0.46)	0.49*** (0.11)	-0.04 (0.15)
Bi. Trade Shares (S)	4.77* (2.11)	0.10 (0.12)	-1.39* (0.65)
S. Value Added (A)	0.34** (0.11)	-0.04 (0.19)	-0.10** (0.04)
S. Value Added (M)	0.42*** (0.12)	0.27* (0.14)	-0.16*** (0.05)
S. Value Added (S)	-0.04 (0.31)	0.11 (0.4)	-0.51*** (0.09)
Urban Pop. Growth	-0.21 (0.21)	-0.20 (0.19)	0.24** (0.08)
Rural Pop. Growth	0.003 (0.02)	0.15 (0.16)	0.004 (0.01)
Total Land Area	0.25*** (0.06)	-0.51 (0.95)	0.008 (0.02)
GDP Growth	0.01 (0.03)	0.36*** (0.08)	0.007 (0.01)
Food Prod. Index	-0.65*** (0.17)	-0.32 (0.22)	0.52*** (0.06)
Crop Prod. Index	0.42** (0.14)	0.90*** (0.26)	-0.22*** (0.05)
Agricultural Land	0.61*** (0.13)	1.01* (0.55)	-0.19** (0.05)
Number of Groups = 33 Number of Time Periods = 8-17 Number of Observations = 464 *** 1% significance level; ** 5% significance level; * 10% significance level; constant excluded.			

Table 16: Regression Results Using Robust Covariance Matrix Estimators for Employment in All Sectors, 1995-2011

While the regression coefficients do not change when calculating the White standard errors and changing the specification of the covariance matrix of the estimated coefficients, we can see that the standard errors do change. Considering the standard errors determine the variability (or confidence interval) of a coefficient estimate, larger standard errors increase the uncertainty and small standard errors decrease the uncertainty of the tested regression coefficient estimate. The standard errors in the RE robust regressions are, for the most part, do not deviate from the values in their respective RE regressions in Sections 4.2, 4.3, and 4.4. Some standard errors are larger in magnitude than their counterparts, and some smaller, but the higher standard errors may not be indicative of less precise estimates. They could be indicative of the estimates being close to zero.

It is important to note that the relatively small magnitude of the regression results within the study do not suggest a small total effect of each variable on employment in each sector. For example, a 0.42% increase in the share of total employment in agriculture, that is 10%, results in a share of 14.2% of total employment in a country dedicated to agriculture. If employment in agriculture was $22.2 million[5], then the 4.2% increase is equivalent to an increase in 9.32 million jobs in agriculture. Thus, although the estimates are quite small in the context of the regression results, they have potential sizeable consequences for labor statistics and economic growth.

Looking back at the key conclusions of the previous literature in Section 2, it is clear that with the utilization of robust covariance matrix estimation (to account for heteroskedasticity and serial correlation in the data), the three RE regression models for employment give us an appropriate fit to test the effect of changes in trade measures, comparative advantage measures, and sectoral productivity on sectoral employment. Hence, these models can reasonably tell us how these factors affect structural transformation through changes in the share of total employment across sectors. In this section of the analysis attempted to empirically verify the theoretical conclusions of previous studies, and the estimation results demonstrate that this research has been reasonably successful. I summarize the conclusions of this data analysis in relation to these conclusions as follows:

1) Empirically, this panel data analysis has demonstrated that the Ricardian trade model - using data computed from theoretical derivations – through the use of a robust covariance matrix RE model, is appropriate in the study of the effects of comparative advantage, trade, and sectoral productivities on structural transformation.
2) Bilateral trade shares (the trade measure) have large, consistent, and statistically significant effects on employment in all three sectors, with a positive or negative employment elasticity of approximately 0.04-4.77% for every 1% increase in a sector's bilateral trade share. Hence trade is clearly important in structural change across all three sectors.
3) Sectoral labor productivity, measured through share of value added in gross output (the productivity measure) per sector has consistent statistically significant results, although its corresponding regression coefficients have a smaller magnitude than those of the trade

[5] As it was in 2019, https://www.ers.usda.gov/data-products/ag-and-food-statistics-charting-the-essentials/

variables above. Hence sector labor productivity, which reflects technological progress and labor efficiency within the sector, is as well important for structural change.

4) Geographical and country-specific factors which represent measures of comparative advantage also affect shifts in sectoral employment. While population measures do not appear to have a statistically significant effect, geographical variables such as the food and crop production indices, and total land area have potential to affect employment all three sectors (but most noticeably employment in agriculture, as expected) .

In terms of the "big picture," these results do have several economic implications. If we consider structural change (particularly the shifting of labor into mostly the service sector) synonymous with development and economic growth, then focusing on improving trade and productivity in all three sectors, and further developing agricultural and crop capacity appear to be the best decisions to improve economic outcomes early on. In the latter stages of structural transformation, the continued development of cities and the urban population appear to be the preferable strategy. Although these conclusions are evident in the data analysis above, it still remains to be seen if the robust results are consistent within the three country income groups represented in my sample. In the following section, I will apply the same regression to the high income country group.

5. Analysis of High Income Countries

In this section, I will test the robust regression model from the empirical analysis results in Section 4 on the group of countries that are classified by the World Bank as "high income." These include all of the countries in the sample excluding Brazil, China, Mexico, Indonesia, and India. Thus, I subset my original sample of 33 countries into two groups – high income and all others – and applied the RE regression model to this subset sample. The intention with the analysis in this section was to test the robust regression model on all three income groups present within the sample: high income, upper-middle income, and lower-middle income. However, due to the composition of the sample, the upper-middle income group was only comprised of 3 countries and the lower-middle income group of 2 countries. Hence, I was unable to use these samples for three separate regressions. Therefore, it is important in the future to widen the sample to include more middle income and low income countries in order to be able to test these effects on those groups separately.

Applying the robust regression to this group of high income countries can still provide us with more information on how trade, productivity, and comparative advantage affect structural transformation within rich countries. Table 17 on the next page presents these regression results.

While for the most part the regression results do not vary significantly from the robust regression results (in terms of sign and magnitude) in Section 4.4, there are some notable changes that should be discussed. First is the role of urban population growth in agriculture; the coefficient estimate in these regression results is highly statistically significant, and expectedly indicates that a 1% increase in the urban population growth rate results in a 0.77% decrease in the share of total employment in agriculture. In addition, an increase in the food production index does not cause a decrease in agricultural employment compared to the model regressed on the whole sample. This indicates that increased food production does not have a negative effect on agricultural employment in richer countries.

In the regression for employment in industry (manufacturing), I now find that increased productivity in the industrial sector (in the form of a higher share of value added) has a negative effect on industrial labor. Specifically, a 1% increase in the share of value added in manufacturing results in a 0.31% decrease in the labor share dedicated to industry. Because rich countries have transitioned, or are in the process of transitioning, from an industry-based to a service-based economy, additional productivity in industry (particularly with modernization of industrial technology) may not attract additional industrial labor or divert workers from entering the service sector. The regression results from employment in the service sector is unsurprisingly (relatively) unchanged, as high income countries have more service labor as a share of total employment than lower income countries do.

Variable	***RE: Agriculture (A)***	***RE: Industry (M)***	***RE: Services (S)***
Bi. Trade Shares (A)	1.54** (0.56)	0.21 (0.13)	-0.22 (0.18)
Bi. Trade Shares (M)	1.57*** (0.36)	0.49*** (0.12)	-0.13 (0.12)
Bi. Trade Shares (S)	2.96 (1.94)	0.14 (0.12)	-1.41* (0.64)
S. Value Added (A)	0.27* (0.12)	-0.15 (0.13)	-0.10** (0.04)
S. Value Added (M)	0.35* (0.15)	-0.31* (0.12)	-0.08* (0.05)
S. Value Added (S)	-0.16 (0.29)	0.08 (0.21)	-0.64*** (0.09)
Urban Pop. Growth	-0.77*** (0.21)	-0.07 (0.14)	0.31*** (0.07)
Rural Pop. Growth	-0.01 (0.03)	0.19 (0.14)	0.01 (0.01)
Total Land Area	0.06 (0.04)	-0.89 (0.84)	0.02* (0.01)
GDP Growth	0.08 (0.06)	0.01 (0.09)	0.03 (0.02)
Food Prod. Index	0.04 (0.22)	-0.01 (0.22)	0.24** (0.07)
Crop Prod. Index	0.24* (0.13)	0.69** (0.26)	-0.13** (0.04)
Agricultural Land	0.45*** (0.11)	0.43 (0.54)	-0.13*** (0.04)
Number of Groups = 33; Number of Time Periods = 8-17; Number of Observations = 464 *** 1% significance level; ** 5% significance level; * 10% significance level; constant excluded.			

Table 17: Regression Results Using Robust Covariance Matrix Estimators for Employment in All Sectors for High Income Countries, 1995-2011

To conclude, we can see that the robust regression model from Section 4.4 is a good fit for the sample of high income countries, and that the results generally do not change much from their respective counterparts in Section 4. The results continue to be consistent with theory, and bolster the conclusions previous authors in this field of study have arrived at.

6. CONCLUSIONS

In this research study, I attempted to answer the question whether measures of trade, comparative advantage, and sector specific productivities have an effect on the rate and incidence of structural transformation, where structural transformation is represented by employment in the agricultural, industrial, and service sectors.

To this end, I gathered data from the WIOD, GGDC-10 database, the PWT database, and the World Bank Development Indicator database on a sample of 39 countries from the years 1995-2011. Extracting the computed values by using Sposi's (2019) and Swiecki's (2017) theoretical 3-sector Ricardian trade model framework, and considering the structure of the data, I implemented several regression models to the panel data. This allowed me to measure the effects of independent variables on employment in each sector by following and comparing the same country over a period of 8- 17 years.

Beginning with the naïve benchmark Pooled OLS regression, I followed with FE and RE model estimations, and verified the role that trade measures, comparative advantage measures, and sectoral labor productivity measures play in patterns of structural change. The FE and RE estimations were consistent in patterns of magnitude, sign, and statistically significant. The magnitudes of the coefficient estimates were large and the majority were statistically significant at the 1% or 5% significance level. Then running diagnostic tests on the chosen RE model, I was able to arrive at a more comprehensive, clear, and accurate model utilizing robust covariance matrix estimators. I then subset the main dataset into two different income level groups of countries. Applying the robust RE model to the high income group, I presented evidence that the robust RE model is an appropriate fit for the (current) higher income groups.

As demonstrated in my robust RE regression results, structural change is primarily affected by the changes bilateral trade shares of each of its respective sector, and the level of productivity in each sector through the share of value added in gross output in its respective sector (denoted by statistically significant coefficient estimates). Shares in employment in each sector are as well affected in very clear ways by comparative advantage measures, such as how share of employment in agriculture increases with positive changes in land area, crop, or food production. Similarly, the share of employment in services is positively affected by increases in the urban population within a country. These patterns are equally evident when applying the robust RE to the high income country group in Section 5.

Although this study allows us to see how certain factors affect employment moving out of, or into, one sector, I will avoid giving any specific policy recommendations. This is mainly due to the fact that the recommendation would depend on which sector is most productive for a specific economy. For example, if Mexico has a comparative advantage in agriculture and manufacturing, then I would encourage an increased share of bilateral trade in agriculture as well as more focus on developing the agricultural sector. I could equally encourage increased shares of bilateral trade in manufacturing or industry. Since these recommendations depends on the productive and resource capacities of each country, I cannot give general policy recommendations that would be appropriate for all of the countries in the same (and especially not those in the population that were not in the same). That being said, the results are easy to interpret to arrive at any recommendations so long as countries have a specific sector they wish to develop; in all, they should focus on comparative advantage measures as well as sectoral productivity instead of increased or decreased sector-specific trade. In addition, it gives us insight into the detailed effects certain factors have on sector-specific employment, which is key to understanding the drivers of structural transformation.

The next steps in continuing this research include expanding the time dimension to improve the scope, accuracy, and forecasting ability of the RE model, as well as introducing other measures of comparative advantage and international trade that may be more useful in building a more accurate and appropriate model. It is also important to apply this model to a more representative and balanced sample of countries, both in terms of income groups and geographic regions. Until then, this research has shown that – using this data – the theoretical conclusions of past papers have empirical verification, and the sensitivity of structural change to economic factors is more pronounced than speculated.