Connor Martins

Professor Dave

EC 483

5 May 2023

Quantifying an Augmented Gravity Model of International Trade Among Professional European Soccer Leagues

Abstract

Gravity models are a popular method used to predict the bilateral trade volume between country pairs and has been studied extensively in commodity markets and migration. The models explore the effects of distance between country pairs, their relative economic sizes, and levels of cultural proximity on the bilateral trade flows between them. This research attempts to extend the model onto the market for professional European soccer players in ten of the continent's elite competitions through a league-level analysis. The empirical estimations were conducted using a two-way fixed effects model and find significant effects of gravitational variables including financial league size, common language and border variables, and transportation costs on bilateral trade volumes.

I. Introduction

Professional European soccer grew dramatically from a financial perspective over the last four decades, particularly in England, Spain, and Italy. Increasing ticket and broadcast revenues drew billions in investment in equity stakes of Europe's biggest clubs, who in turn splash cash on players in the hopes of qualifying for extremely lucrative tournaments like the UEFA Champions

League, the annual prize pot of which was 2.06 billion Euros in 2021. With the returns on competitive success higher than ever in Europe’s most elite leagues, international player transfers are becoming costlier and the subject of increased scrutiny and regulation. Transfer fees like that of Neymar Jr., a Brazilian player who moved from Spanish outfit FC Barcelona to Paris Saint-Germain, a French club, in 2017, have exceeded 200 million Euros. While some argue that such fees represent market distortion by state-owned clubs like Paris Saint-Germain and Manchester City, there is no debate over the dramatically increasing global and financial importance of the sport.

The research attempts to quantify an augmented gravity model of trade on the market for European professional soccer players among ten of its most elite leagues, including the ridiculously lucrative English Premier League, which boasts sky-high player salaries and broadcasting rights contracts. Quantifying the model will determine if the market for Europe’s elite soccer players can be understood and predicted through variables like trade distance, cultural similarities, and the market values of the leagues involved in the transactions, similar to how conventional trade gravity models define commodity trade relationships between country pairs. Specifically, the gravity model will attempt to quantify the level of bilateral trade volume between the ten countries over the years between 2012 and 2020, a period over which European soccer experienced tremendous financial growth, particularly in broadcasting contracts and foreign investment.

Understanding the market mechanisms behind player transfers and the key factors that inhibit or promote international trade between leagues is useful for industry stakeholders and regulators seeking to retain competitive balance seen as key to the long-term success of the sport. Furthermore, it is worth considering whether the rapid integration of enhanced data analytics into the sport along with telecommunications developments have put downward pressure on the traditional “distance effect” on international trade. In other words, the effect of trade distance and other gravity-related factors like sharing a border and a common language may have been impacted by technologies that make transfer-related activities like player scouting and fee negotiations faster and cheaper. The research will attempt to determine whether geographical and cultural proximity, positional differences in player profiles, and the relative sizes of professional soccer leagues have significant impacts on the bilateral trade volume between them.

II. A Review of the Relevant Literature

The gravity model of physics, which relates the gravitational pull between two objects to their masses and the distance between them, became an important method of predicting international trade flows in economics, starting as early as 1962 with Jeffrey Bergstrand's seminal paper. OLS regressions and several other estimation methods are used for gravity modeling with bilateral trade flows as the outcome. The independent variables almost always include country pair values of population, GDP, and the great circle distance between the capitals of the countries being studied. It is also popular to use some kind of dummy variables that reflect cultural similarities between trading partners, such as common language, border, or colonial ties dummies. The outcome and the continuous independent variables are usually logged in gravity model estimations. Distance is expected to have negative impacts on trade flows, whereas population and GDP are expected to have the opposite effect, and the literature confirms this with impressive accuracy. Indeed, in his 2010 paper for the National Bureau of Economics Research, James Anderson called the gravity model "one of the most successful empirical models in economics". Its high predicting power and growth in popularity has paved the way for the study of the marginal effects on trade flows of other variables, such as trade restrictions or agreements. Gravity models are not restricted to the study of the flow of commodities. In 2006, Lewer and van den Berg quantified a gravity model of immigration using panel data from sixteen OECD countries using a similar mathematical structure as employed by commodity gravity models. They also tested the effects of the number of people from the origin country already living in the destination country, which is expected to lower the costs associated with adapting to life in a new country. Lewer and van den Berg then compared the gravity model of immigration to one of trade, finding the same coefficient signs and levels of significance. Only the common border dummy was insignificant, which is an important takeaway for my augmented model as it suggests that the flow of people across borders is more rigid than the flow of commodities.

The *ad hoc* nature of Bergstrand's early gravity foundations have been criticized and revised, but beyond these issues arose a particularly interesting phenomenon of the gravity model: the distance elasticity puzzle. This paradox arose from the estimation of separate cross-sectional gravity models at different points in time, which seemed to generate distance elasticities that rose in absolute value. In other words, according to the gravity model literature, distance appeared to be having an increased (in this case, “more negative”) impact on trade flows over time. This contrasts with the overwhelmingly popular notion that globalization has reduced such international trade inhibitions. Brun, Carrere, Guillamont, and de Melo set out in 2005 to explain this issue, pointing towards potential causes such as “misspecification” in transport costs functions and omitted variable bias (OVB). Brun et al. used a more thorough treatment of the transport costs function, using indexes that accounted for oil prices and levels of infrastructure among other factors. Their statistical approach was also unique, using a random-effects panel procedure to combat endogeneity issues through the incorporation of Hausman Taylor estimators. The improved model, which also included a quadratic distance term to capture a potential inflection point, indicated an eleven percent decline in distance elasticities over the thirty-five year period studied. However, after controlling for income of the countries in the data, the puzzle was found to persist in bilateral trade between low-income countries, suggesting that poorer nations have not necessarily reaped the benefits of globalization on transport costs.

An important application of the gravity model is to analyze the consequences of trade barriers and the benefits of trade agreements, and has therefore been well-studied. Wall 2000 investigated the costs of American protectionism in the 1990s using a fixed effects gravity model with a “trade policy index” independent variable, finding that it had a significant and negative impact on trade flows when removing statistical restrictions on the intercept. He then extrapolated the results to find that U.S. protectionism caused around a 26.2 percent loss of “potential merchandise exports to non-NAFTA countries”, relative to having a more open trade policy with said countries in 1996. Bergstrand 2014 attempted something similar, examining the benefits of “economic integration agreements” (EIAs) on trade flows, finding positive effects after controlling for endogeneity. Brexit in particular has become a subject of the gravity model literature due to its imposition of trade restrictions. A study by Karlsson, Melin, and Cullinane in 2018 drafted a “pessimistic” Brexit scenario (it had not yet been passed at the time of publishing) that used projected U.K. GDP reductions and tariff rate increases, and the model consequently predicted that Germany would export 102,000 fewer cars to the U.K. relative to the “no Brexit” scenario, showcasing the powerful ability of the gravity model to predict the marginal effects of such trade policies.

My contribution is significant because of the increasing economic presence of professional European soccer and because I could not find any major studies that attempted to model transfer flows using gravity. As soccer leagues attract more investment into broadcasting rights, commercial sponsorships, and the clubs themselves, it is useful to understand the nature of and ideally predict the trade flows between leagues. An interesting application of the model would be to investigate the effect of Brexit, which now requires players from foreign countries to apply for British work permits if bought by an English Premier League (EPL) club, on trade flows between the EPL and the other nine leagues in the data set. Additionally, I am interested in investigating whether distance has indeed had a declining impact on trade between the leagues as discussed by previous studies, especially since the process of scouting players has been made easier through the internet and enhanced analytics products (such as those provided by companies like Oracle Cloud). A well-fitting gravity model would be useful to various stakeholders including league organizers looking to improve competitive balance, regulators investigating the nature of modern player transfers, and agents looking to maximize the financial return of their negotiations with interested buyers. Soccer clubs themselves would likely benefit most from such a model as they would be able to better understand the underlying reasons behind the way international transfer business is conducted and how to use them to their advantage.

III. Conceptual Framework

The augmented gravity model of trade is based on the physics equation that computes the gravitational pull between two objects as a function of their mass, the distance between them, and the gravitational constant. The relationship is demonstrated in (1).

$$F_g = G\frac{m_1 m_2}{r^2} \tag{1}$$

By (1), the gravitational pull is inversely proportional to the distance between the objects and proportional to their relative sizes. Therefore, a basic gravity model of trade for a country trading pair *i* and *j* would use their respective gross domestic products at their economic "mass" and the great circle distance between them as the analogous variables of the physics equation. The model is shown as a functional relationship in (2).

$$T_{ij} = \frac{Y_i \, Y_j}{D_{ij}} \tag{2}$$

In (2), the respective gross domestic products of countries *i* and *j* are multiplied in the numerator, and the denominator consists of the great circle distance between them. The function theoretically yields the bilateral trade volume, denoted as T_{ij}. In a functional, linear regression form, the gravity model is estimated as shown in (3).

$$T_{ij} = \beta_0 + \beta_1 \ln(Y_i \ln Y_j) + \beta_2 D_{ij} + \varepsilon_t \tag{3}$$

In (3), the values of β_1 and β_1 can be interpreted as the marginal effects of the product of the countries' mass and the distance between them on their bilateral trade volume, assuming some error term ε_t. If the gravity fits the data of interest, the first coefficient should be positive—larger nations are more likely to trade with each other. Additionally, the second coefficient should be negative because greater distance between countries reduces the likelihood of trade occurring, a concept related to the increased transaction costs inherently implied by distance. While the basic elements of this theory suggest expected directions of the marginal effects, the augmented gravity model will incorporate key variables of interest beyond economic size and simple trade distance. However, it provides a necessary benchmark for evaluating the output of various estimation methods and assessing the applicability of gravity to this less conventional international trade market, particularly because it involves the transfer of human beings between countries. For the dataset of interest, this basic model will be adapted and augmented to reflect the less direct measurement methods of the economic and physical characteristics of the European soccer leagues involved in international player transfers. The estimation of the model will test the reasonableness of the essential gravity equation assumptions outlined in this section.

IV. Data and Variables

The key dependent variable of interest in the research is the bilateral trade flow between the country pairs in the dataset over the period between the start of 2012 and the end of 2020. It is worth noting that in European soccer, players can only be transferred between teams during two specific windows: once in the summer between the months of June and August, and once in the winter during the month of January. Therefore, the sum of transfer fees from the original dataset are grouped not just by year, but also by the transfer window in which they occurred. Furthermore, the transfer fees are aggregated by the sport-related positions of the players involved in the transfers, which are categorized as goalkeeper, midfielder, defender, or attacker. They were grouped by position because the market values and transfer fees of players across different positions vary greatly, especially in more competitive leagues like the English Premier League, which pay out enormous sums for attacking players. Goalkeepers also tend to be particular expensive due to their rarity. Controlling for position will help to eliminate potential biases in the estimation results. Finally, to avoid symmetrical issues in the computation of bilateral trade volume, they are also grouped by a dummy variable indicating whether the transfer fee was for a player purchase or a player sale. The final result of the data aggregation is a sum of transfer income and expenditure between unique country pairs grouped by year, position, and transfer window, called bilateral trade volume. Bilateral trade volume is measured in Euros and is a nominal variable. The original dataset consisted of almost 45,000 player transfer data points compiled through a web-scraping program that extracted data from Transfermarkt.com, the most popular public database of international soccer player transfer values, fees, and other relevant information. Aggregating the data into bilateral trade volumes for unique country pairs using the discussed method resulted in a total of 7,124 observations in the final dataset.

Key independent variables include the product of the economic sizes of the two leagues involved in a given transaction. The economic size was determined by the summation of the market values of each of the clubs in the league in that particular year, measured in nominal Euros. Each of the ten leagues has between sixteen and twenty participating clubs every season, though the actual clubs change due to the competitive relegation and promotion structures with non-top-flite divisions in the country. This data is also available on Transfermarkt.com and was manually extracted for the research. The player position dummies were derived from information in the original, web-scraped dataset. Consistent with the conceptual framework, the natural log of the product of the two league sizes was taken before being used in any model estimations.

Another key independent variable is the great circle distance between the geographic center of the two trading countries, which is computed using the Haversine distance formula and is measured in kilometers. Common border and language dummy variables were constructed based on the ten countries in the dataset: England, Spain, Germany, Italy, France, Portugal, the Netherlands, Russia, Belgium, and Scotland, and used as regressors in the models to measure cultural and geographical proximities. Also, the average price of crude oil in Europe was used as an independent variable to reflect transportation costs. Average July and January prices for each year in the dataset were used and converted from USD per barrel to nominal Euros per barrel based on average spot rates reported by OFX.com, a popular currency converter. There were 7,124 observations of each variable except for the mutually exclusive dummy variables, whose relative proportions and counts are summarized in Table 1. Figures 1 and 2 illustrate the distributions of two of the key variables discussed.

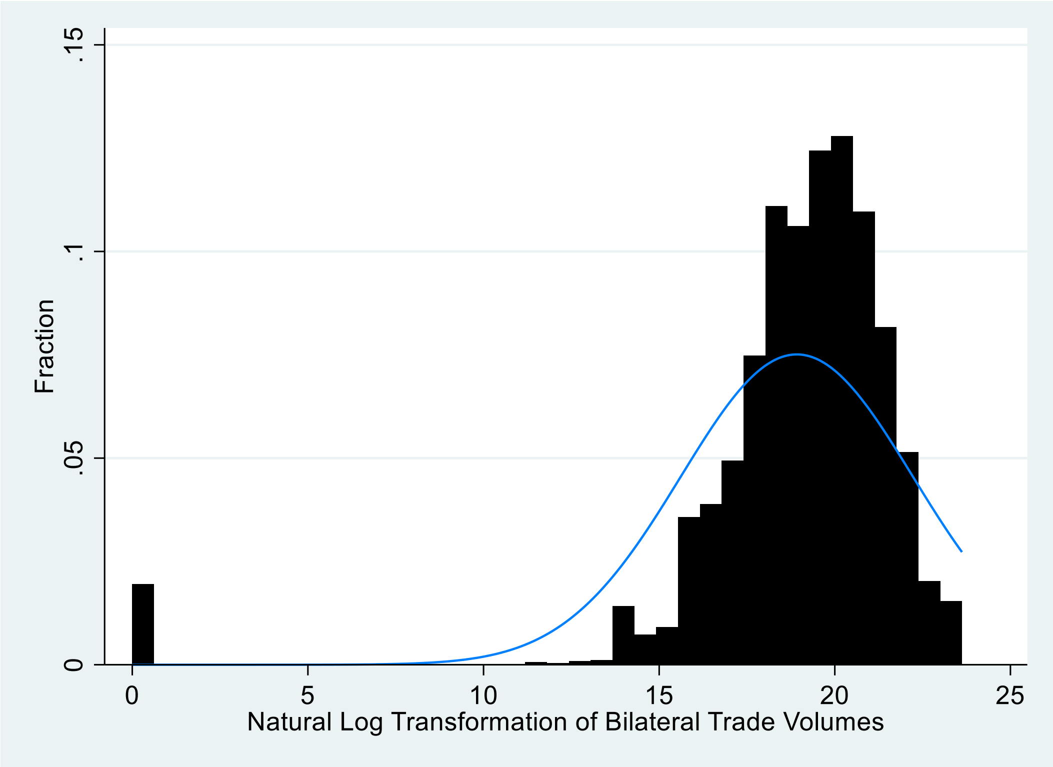

Figure 1: Histogram of the Natural Log Transformation of Bilateral Trade Volume

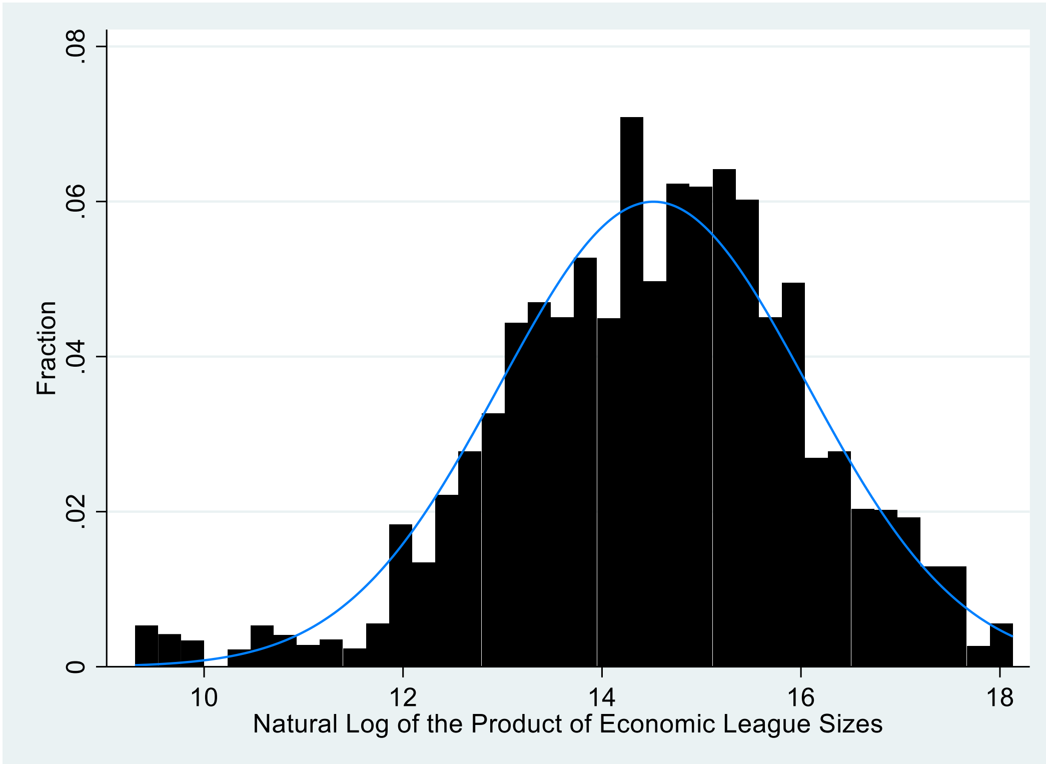

Figure 2: Histogram of the Natural Log of the Product of Economic League Sizes

Dummy Variable Name	Mean	Standard Deviation	Observation Count When Dummy = 1
Common Language Dummy	0.345	0.475	2,459
Common Border Dummy	0.376	0.484	2,676
Purchase (1) or Sale (0) Dummy	0.482	0.5	3,432
Goalkeeper Dummy	0.133	0.339	950
Defender Dummy	0.291	0.454	2,076
Midfielder Dummy	0.245	0.43	1,745
Attacker Dummy	0.33	0.47	2,353
Summer (1) or Winter (0) Window Dummy	0.648	0.478	4,613

Table 1: Gravity Dummy Variable Descriptive Statistics

V. Empirical Methodology

The empirical methods used were derived by extending the basic gravity model of trade discussed in the conceptual framework onto several other relevant variables that capture potential frictions on or promoters of international trade between leagues. The basic functional representation of the key variables is shown in (4).

$$\ln(T_{ijt}) = \beta_0 + \beta_1 \ln(Y_{it} \ln Y_{jt}) + \beta_2 DIST_{ij} + \beta_3 LANG_{ij} + \beta_4 CONTIG_{ij} + \beta_5 GK + \beta_6 DEF + \beta_7 MID + \beta_8 ATT + \beta_9 SUMMER_t + \beta_{10} BRENT_t + \varepsilon_t \quad (4)$$

Where:

- T_{ijt} is the bilateral trade volume between leagues *i* and *j* in year *t*
- Y_{it} and Y_{jt} are the summations of the market values of the clubs in leagues *i* and *j* in year *t*
- $DIST_{ij}$ is the great circle distance between the geographic centers of the countries in which leagues *I* and *j* are located
- $LANG_{ij}$ is a dummy variable that is equal to 1 if leagues *i* and *j* are located in countries with a common language
- $CONTIG_{ij}$ is a dummy variable that is equal to 1 if leagues *i* and *j* are located in countries with a common border
- *GK*, *DEF*, *MID*, and *ATT* are dummy variables equal to 1 if the bilateral transfer volume is of goalkeepers, defenders, midfielders, or attackers, respectively
- $SUMMER_t$ is a dummy variable equal to 1 if the bilateral transfer volume is of transfers that took place in the summer transfer window in year *t*
- $BRENT_t$ is a continuous variable that reflects the average price of crude oil in Europe in nominal Euros in the summer or winter (depending on the value of $SUMMER_t$) of year *t*

The inclusion of the four player position dummy variables builds in a control for potential differences between the market values of players in different positions. Fitting these dummy variables before estimating any models eliminates potential unobservable differences at the positional level and reduces bias in all estimated effects. All the variables in (4), the main equation, are either continuous or dummy variables. As is typical of the trade literature, the natural log of the product of the economic sizes of leagues *i* and *j* was taken and will improve the interpretability of its coefficient. Its coefficient will be interpreted as an elasticity since the outcome of the equation is also in natural log form.

Applying the conceptual framework, the expected signs of each coefficient in (4) were hypothesized. β_1 should be positive, as the product of the relative economic masses of the leagues should increase trade between them. β_2 should be negative due to the hypothesized inverse relationship between trade flows and physical distance between league pairs. β_3 should be positive, as increased levels of cultural proximity are believed to have positive influences on trade volume between countries in the gravity model literature. β_4 should also be positive, as having a shared border should theoretically promote trade by acting as a proxy for lower transaction costs, cultural similarities, and lower transportation costs in the scouting of players. For example, the likelihood of a player in one league joining a club in another is expected to increase if the costs associated with moving to the new country are limited by a shared border, all else equal (e.g. salaries, career aspirations, etc.). The coefficients on the player position dummies have unknown expected directional effects and will be evaluated after various model estimations. β_9 is expected to have a positive effect on bilateral trade volume simply because the summer transfer window is two months longer than the winter window and is a far more popular time for clubs to invest or divest in their squads. Finally, β_{10} is expected to be negative, as average oil prices should proxy transportation costs and should inhibit international trade, a hypothesis consistent with the gravity model of trade literature.

It is highly likely that there will be unobservable differences across time and across leagues that could bias the magnitudes or directions of these coefficients, and that issue will be tested for and eliminated through statistical methods during estimation. A player-level model will also be estimated that attempts to quantify a potential relationship between the difference a player’s transfer fee and his market value at the time of the transfer and the independent variables used in (4). It is important to note that this is not a direct gravity model estimation and is secondary to the main equation, (4). However, this equation could determine investment inefficiencies related to the underlying gravitational mechanisms that dictate transfer flows studied in the main empirical framework and could therefore be a valuable byproduct of the research. The framework of the secondary equation is laid out in (5).

$$VALUE\_DIFF_{pijt} = \beta_0 + \beta_1 \ln(Y_{it} \ln Y_{jt}) + \beta_2 DIST_{ij} + \beta_3 LANG_{ij} + \beta_4 CONTIG_{ij} + \beta_5 GK + \beta_6 DEF + \beta_7 MID + \beta_8 ATT + \beta_9 SUMMER_t + \beta_{10} BRENT_t + \varepsilon_t$$

Where $VALUE\_DIFF_{pijt}$ is the difference between a player *p*'s market value in nominal Euros and the transfer fee paid for him in year *t*, where the player is being transferred from league *i* to league *j*. For this variable, a positive value represents a bargain for the buying league *i*—in other words, the team in league *i* paid less than the market value for the player *p*, and vise versa for a negative value. All independent variables in (5) are the same as those used in (4). While this method is generally understudied and the effects of most of the coefficients are unknown, reasoning leads me to expect that β_2 will be negative, as distance increases the difficulty of accurately scouting and financially valuing a player before offering a bid to the dealing club. Furthermore, β_8 is expected to be negative, as attacking players in European soccer are typically subject to transfer fees that far exceed their market value.

A variety of estimation methods will be used to quantify both (4) and (5), the results of which will be reported in Section VI.

VI. Results

Several panel data models were employed in the quantifying of the augmented gravity model. A naïve OLS regression was estimated and revealed highly significant gravity regressors. In particular, the product of the economic league size variable was significantly negative as predicted. However, more robust estimation methods were used to better account for potential issues like selection bias across leagues and over time.

	Naïve OLS			League-Level FE Model		
	Coefficient	Std. Error	p-value	Coefficient	Std. Error	p-value
$\ln(T_{ijt})$						
$\ln(Y_{it} \ln Y_{jt})$	1.1314	0.0211	0.000	1.1117	0.0292	0.000
$DIST_{ij}$	-0.0001	0.0000	0.000	-0.0001	0.0000	0.000
$LANG_{ij}$	1.4225	0.0882	0.000	1.3856	0.0898	0.000
$CONTIG_{ij}$	0.5887	0.0814	0.000	0.5324	0.0861	0.000
GK	0.2408	0.1041	0.021	0.3566	0.0979	0.000
DEF	0.0230	0.0831	0.782	0.1498	0.0752	0.047
MID	0.0000	(omitted)		0.1036	0.0791	0.190
ATT	-0.1088	0.0810	0.179	0.0000	(omitted)	
$SUMMER_t$	0.4331	0.0643	0.000	0.4197	0.0629	0.000
$BRENT_t$	-0.0437	0.0018	0.000	-0.0483	0.0018	0.000
Constant	4.25017	0.36785	0.000			

Table 2: Estimation Results of the Naïve OLS and League-Level FE Models

Table 2 demonstrates the results of the OLS regression, which estimated highly significant gravity regressor coefficients of the expected signs based on the conceptual framework. It was concluded that the model needed to undergo several robustness checks before moving on to more sophisticated estimations. A Breusch-Pagan/Cook-Weisberg test was conducted to look for heteroskedasticity in the dataset, producing a p-value of 0.000 and strong evidence that this is an issue with the data. The OLS regression was re-estimated with robust standard errors, and the results are shown in Table 4. Fortunately, generating robust standard errors did not affect the significance of any of the relevant gravity regressors. The data was also tested for multicollinearity using a pairwise collinearity matrix, shown in Table 5, and the computation of the Variance Inflation Factors (VIF) among the variables. No VIF's exceeded 2.0, which indicated that multicollinearity among the regressors is not an issue for the dataset. The same conclusion was reached from the collinearity matrix, which also produced no concerningly high values. Finally, to test for autocorrelation, a Durbin-Watson was computed. The value was roughly 1.55, which indicates some degree of positive autocorrelation among the residuals of the OLS regression. A Prais-Winsten procedure was used to re-estimate the OLS regression to correct for serial autocorrelation. The re-calculated Durbin-Watson statistic after this procedure improved to just over 2.0 without losing significance or having material impacts on the gravity coefficients. After a series of robustness tests, further estimation was conducted to improve the gravity modeling.

	$\ln(Y_{it} \ln Y_{jt})$	$DIST_{ij}$	$BRENT_t$
$\ln(Y_{it} \ln Y_{jt})$	1		
$DIST_{ij}$	-0.0138	1	
$BRENT_t$	-0.1928	-0.0022	1

Table 4: Pairwise Collinearity Matrix of Continuous Gravity Regressors

	Time FE Model			League and Time TWFE			Random Effects Model		
	Coeff.	Std. Error	p-value	Coeff.	Std. Error	p-value	Coeff.	Std. Error	p-value
$\ln(T_{ijt})$									
$\ln(Y_{it} \ln Y_{jt})$	1.0808	0.0201	0.000	1.0126	0.0298	0.000	1.0808	0.0201	0.000
$DIST_{ij}$	-0.0001	0.0000	0.000	-0.0001	0.0000	0.000	-0.0001	0.0000	0.000
$LANG_{ij}$	1.3258	0.0799	0.000	1.3342	0.0812	0.000	1.3258	0.0799	0.000
$CONTIG_{ij}$	0.5963	0.0734	0.000	0.5513	0.0778	0.000	0.5963	0.0734	0.000
GK	0.3815	0.0904	0.000	0.3959	0.0884	0.000	0.3815	0.0904	0.000
DEF	0.1669	0.0695	0.016	0.1829	0.0679	0.007	0.1669	0.0695	0.016
MID	0.1450	0.0730	0.047	0.1449	0.0714	0.042	0.1450	0.0730	0.047
ATT	0.0000	(omit)		0.0000	(omit)		0.0000	(omit)	
$SUMMER_t$	0.4446	0.0588	0.000	0.4324	0.0575	0.000	0.4446	0.0588	0.000
$BRENT_t$	-0.038	0.0100	0.000	-0.038	0.0098	0.000	-0.038	0.0100	0.000

Table 3: Estimation Results of the Time-Level FE, Two-Way FE, and Random Effects Models

The league-level FE model was estimated in order to account for unobservable differences across leagues that might have biased the naïve OLS coefficient estimations, called league-level selection bias. League-level selection bias could certainly be an issue with this data, as different leagues employ very different scouting tactics, negotiation styles, and financial methods when acting in the market for professional soccer players. This model attempts to control for these differences and produce less biased results. The league-level FE model found that a one percent increase in the product of the economic sizes of a league pair is associated with a 1.11 percent increase in the predicted bilateral trade flow between the league pair, all else equal. The direction of the coefficient was expected as a result of the theoretical underpinnings of the model, and the magnitude is reasonable. For reference, a 1.11 percent increase in the mean bilateral trade flow in the dataset represents an increase of roughly 5.5 million Euros. The common language dummy variable was also highly positive and significant: when a country pair shares a common language, the league-level FE model expects the predicted bilateral trade flow between the pair to increase by 1.39 percent relative to not sharing a language, all else equal. Finally, the oil price variable also had the expected, negative sign and was significant. A one Euro increase in the oil price is associated, according to this particular model, with a 0.048 percent decrease in the predicted bilateral trade flows between a country pair, all else equal. Finally, the model found that, if the trade between league pairs was of goalkeepers in a given window, the predicted value of the trade volume between them is 0.40 percent higher relative to attacking players. However, the player position dummies are less relevant for the gravity estimation and were largely incorporated to eliminate position selection bias. It is worth noting that adjusting for league-level selection bias reduced the magnitudes of the naïve OLS regression, but did not change the sign or significance of any of the gravity regressors, which is encouraging.

To further test the strength of the model, both a two-way FE model and a random effects (RE) model were estimated. The two-way FE model controlled for both time and league-level selection bias. The previous FE models for each had revealed, through joint significance tests, that league and time-level selection bias were an issue in the OLS model. Indeed, both of the joint significance tests had sufficiently low p-values of 0.000, giving more than enough evidence to reject the null hypothesis that all fixed effects were jointly insignificant. Therefore, the two-way FE model was the next logical step, as it mathematically eliminates both of these types of biases. Estimating the two-way FE model yielded, once again, significant gravity regressors with the expected signs. The league size coefficient decreased slightly (to a 1.02 percent increase in bilateral trade flows), as did the common border and common language dummies, whereas the distance and oil price variables remained the same as in the single FE models. The random effects model predicted similar results, with slightly lower league size and common language dummies but a higher common border dummy than the two-way FE model.

A Hausman Taylor test was conducted to determine whether the TWFE or RE model fit the data better, and the p-value was 0.000, allowing for the null hypothesis of a superior RE model to be rejected. It was concluded that the TWFE model is the better fit for the augmented gravity model since it comprehensively eliminates time and league-level selection biases. Based on the model estimations, the hypotheses regarding the effects of economic league size, distance, common language and borders, and transportation costs held. However, the residual values of the TWFE model were plotted against its predicted values, and there was a clear relationship as shown in Figure 3. This kind of relationship is indicative of heteroskedasticity among the error terms and is a cause for concern about the fit of the TWFE model.

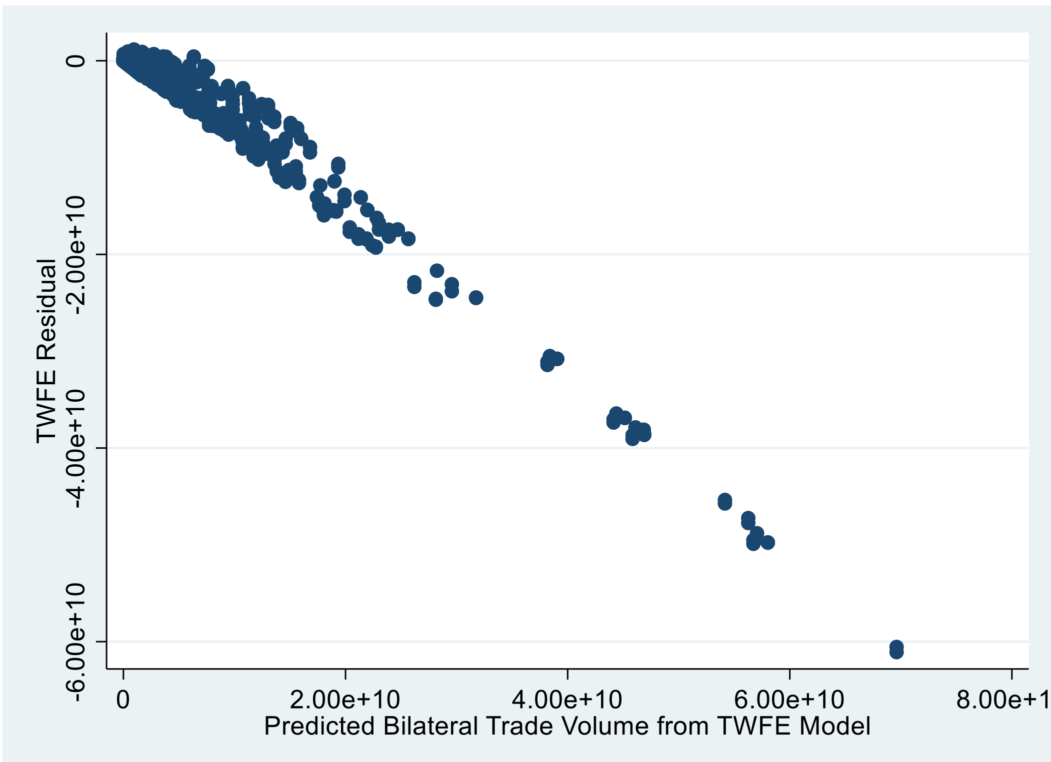

Figure 3: TWFE Residuals vs. its Predictions of Bilateral Trade Flows

To deal with the issue, a TWFE model with heteroskedasticity-robust standard errors was estimated and its results are shown in Table 4. Adjusting for heteroskedasticity in the TWFE standard errors caused the distance and common border dummies, indicating that the previous model was overestimating their effects on bilateral trade volume. Post-adjustment, physical distance was not at all significant in predicting trade flows. The common border dummy and oil price variables lost a level of significance and were only significant at the five and ten percent levels after adjustment. The adjustment had no effect on the magnitude or significance of the league size product and common language variables, which remain significant at all levels.

	TWFE Model with Heteroskedasticity-Robust Standard Errors		
	Coefficient	Robust Std. Error	p-value
$\ln(T_{ijt})$			
$\ln(Y_{it} \ln Y_{jt})$	1.012565	0.149892	0.000
$DIST_{ij}$	-0.000128	0.000119	0.305
$LANG_{ij}$	1.334229	0.341714	0.003
$CONTIG_{ij}$	0.551317	0.233297	0.040
GK	0.395878	0.054984	0.000
DEF	0.182922	0.029342	0.000
MID	0.144885	0.037281	0.003
ATT	0.000000	(omitted)	
$SUMMER_t$	0.432394	0.069573	0.000
$BRENT_t$	-0.038271	0.013984	0.021

Table 4: TWFE Model with Clustered-Robust Standard Errors

The original OLS model suffered from the biases of heteroskedasticity and league and time-level selection bias, adjusted for by the estimation of a TWFE model with clustered-robust standard errors. Figure 4 shows the plot of the observed versus the predicted bilateral trade volumes from the final model and indicates a strong relationship close to a 45-degree line, which would be perfect.

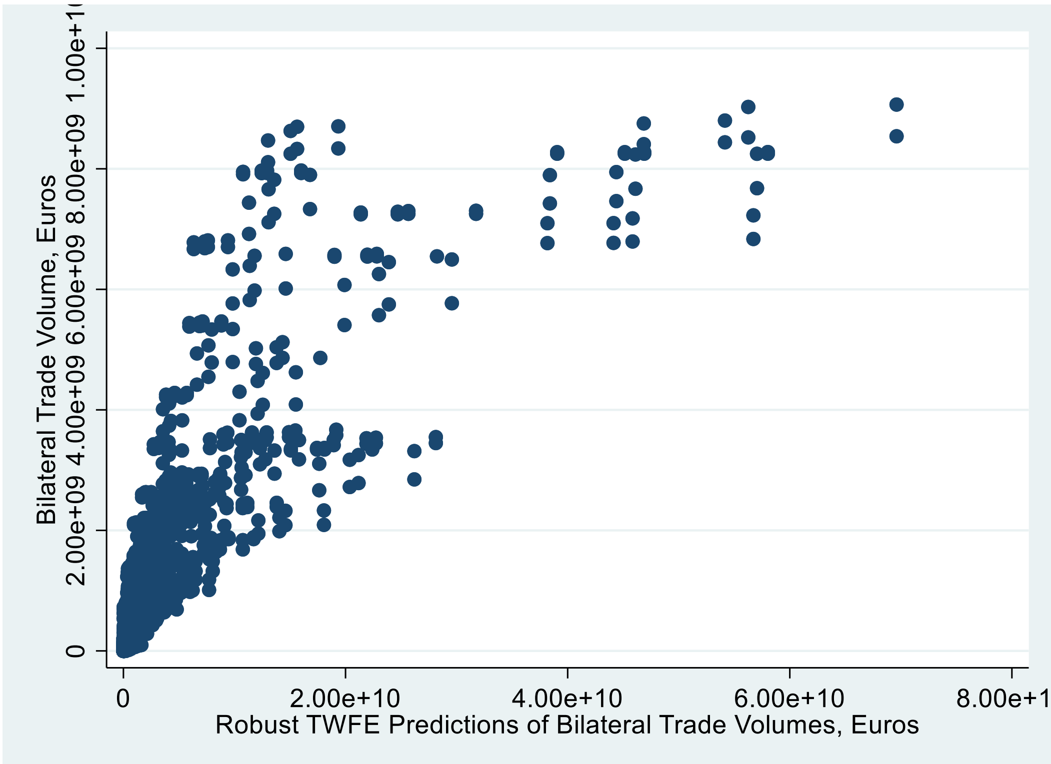

Figure 4: Actual Trade Volumes vs. Robust TWFE Predicted Trade Volumes

	OLS		
	Coefficient	Robust Std. Error	p-value
$VALUE\_DIFF_{pijt}$			
$DIST_{ij}$	107.23930	28.02461	0.000
$\ln(Y_{it} \ln Y_{jt})$	621359.60000	12288.82000	0.000
GK	0.00000	(omitted)	
DEF	341550.90000	76492.86000	0.000
MID	778783.20000	79351.63000	0.000
ATT	1032790.00000	74372.98000	0.000
$LANG_{ij}$	-1753805.00000	80111.13000	0.000
$CONTIG_{ij}$	-54167.56000	77420.52000	0.484
$BRENT_t$	-4849.81700	1187.77500	0.000
Constant	-6070582.00000	223972.30000	0.000

Table 5: Naïve OLS Model for the Player Bargain Variable

Finally, a simple OLS regression of the difference between a player's market value and transfer fee in a given year was regressed against the gravity regressors as the empirical estimation of (5). Because this equation is secondary to the research question, its analysis was kept basic. Interestingly enough, the coefficients were significant, though only some had material magnitudes on market value disparities at the player level. A one-Euro increase in the average price of crude oil is associated with a nearly 5,000 Euro reduction in the "bargain" for a player transaction. Surprisingly, being an attacking player was associated with just over a one million Euro increase in the predicted bargain value relative to goalkeepers, who might therefore be the most expensive types of players on the market. Furthermore, shared borders are associated with a 54,168 Euro decrease in the difference between player value and transfer fee relative to country pairs without a shared border. Finally, having a shared language was associated with a nearly 1.8 million Euro decrease in the bargain for a given player relative to not sharing a language. It is possible that these results indicate that a larger proportion of the "bargains" realized in the market for European soccer players are the trades involving non-goalkeepers and between countries without as much geographical or cultural similarities, potentially due to the sophisticated scouting methods and information sharing technologies required to execute such trades efficiently. More research is needed to determine the true nature or existence of these relationships.

VII. Conclusions

The gravity model of trade can be extended to a league-level analysis of the transfer volumes between Europe's ten biggest professional soccer leagues. After testing various estimation models, the two-way fixed effects model with robust standard errors showed the expected effects and magnitudes of league size, transportation costs, and having a shared border or language. After adjusting for heteroskedasticity, the distance effect on trade volume became completely insignificant, while the predictive power of transportation costs and of the common border dummy decreased slightly. The theoretical underpinnings of the gravity model partly hold in this augmented environment, and while the relationship between trade volume and country pair sizes was as expected, the effect of distance was disappointingly insignificant. Overall, the model fits the data well and shows how fundamental international trade concepts hold in a less conventional economic setting in which the goods traded are not commodities, but professional soccer players.

Additionally, preliminary models were estimated to understand the relationship between gravity regressors and the "bargain" for a professional soccer player. Evidence from the naïve regression suggests that there could be inverse relationships between cultural and geographical proximities between country pairs and the value of the bargain, meaning that clubs could potentially seek more cost-efficient player recruitment by scouting leagues in countries further away and less culturally similar. Understanding comprehensively these relationships require more sophisticated estimation methods and was secondary to the main question of the gravity modeling.

References

References

Anderson, J. (2010). The gravity model. *NBER Working Paper Series*. https://doi.org/10.3386/w16576

Borg, S. (2021, September 18). *Champions League prize money breakdown: How much do the winners get in 2021?* Sporting News. Retrieved May 5, 2023, from https://www.sportingnews.com/us/soccer/news/champions-league-prize-money-2021/5i1j0n2hh9r01ffvgqfsp541l

Bergstrand, J. H. (1985). The gravity equation in international trade: Some microeconomic foundations and empirical evidence. *The Review of Economics and Statistics*, *67*(3), 474. https://doi.org/10.2307/1925976

Bergstrand, J. H., Larch, M., & Yotov, Y. V. (2015). Economic integration agreements, border effects, and distance elasticities in the Gravity Equation. *European Economic Review*, *78*, 307–327. https://doi.org/10.1016/j.euroecorev.2015.06.003

Brun, J.-F. (2005). Has distance died? evidence from a panel gravity model. *The World Bank Economic Review*, *19*(1), 99–120. https://doi.org/10.1093/wber/lhi004

Carrère, C., Mrazova, M., & Neary, J. P. (2020). Gravity without apology: The Science of Elasticities, Distance, and Trade. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3555587

Egger, P. (2001). An econometric view on the estimation of gravity models and the calculation of trade potentials. *WIFO Working Papers*, *141*. https://doi.org/10.1111/1467-9701.00432

Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *The Quarterly Journal of Economics*, *CXXIII*(2), 441–487. https://doi.org/10.3386/w12927

Karlsson, J., Melin, H., & Cullinane, K. (2018). The impact of potential Brexit scenarios on German car exports to the UK: An application of the Gravity Model. *Journal of Shipping and Trade*, *3*(1). https://doi.org/10.1186/s41072-018-0038-x

Lewer, J. J., & Van den Berg, H. (2008). A gravity model of immigration. *Economics Letters*, *99*(1), 164–167. https://doi.org/10.1016/j.econlet.2007.06.019

Owen, P. D., & Owen, C. A. (2019). Simulation Evidence on Herfindahl-Hirschman Indices as Measures of Competitive Balance . *Journal of the Operational Research Society*, *73*(2), 285–300. https://doi.org/10.1080/01605682.2020.1835449

Serlenga, L., & Shin, Y. (2004). Gravity models of intra-EU Trade: Application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics*, *22*(2), 361–381. https://doi.org/10.1002/jae.944

Shepherd, B. (2012, October 1). *The gravity model of international trade : A user guide*. Repository Home. Retrieved March 25, 2023, from https://repository.unescap.org/handle/20.500.12870/128

Szymanski, S. (2000). A market test for discrimination in the English Professional Soccer Leagues. *Journal of Political Economy*, *108*(3), 590–603. https://doi.org/10.1086/262130

Szymanski, S. (2003). The Assessment: The Economics of Sport. *Oxford Review of Economic Policy*, *19*(4), 467–477. https://doi.org/https://www.jstor.org/stable/23606854

Wall, H. J. (2000). Using the gravity model to estimate the costs of protection. *Federal Reserve Bank of St. Louis Review*, *81*(1). https://doi.org/10.20955/r.81.33-40

Zambom-Ferraresi, F., García-Cebrián, L. I., Lera-López, F., & Iráizoz, B. (2015). Performance evaluation in the UEFA Champions League. *Journal of Sports Economics*, *18*(5), 448–470. https://doi.org/10.1177/1527002515588135