INTRODUCING GOOGLE TRENDS IN ANALYSING FINANCIAL INTEGRATION AND EUROPEAN BANK PERFORMANCE

Author:
Alicia Elizabeth Sumampouw
388595

Supervisor:
dr. S.H. Bijkerk

A thesis submitted for the
Bachelor of Science in Economics and Business Economics

August 2017

ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
Department of Economics

Abstract

This thesis analyses the relationship between financial integration and European bank performance in the period 2007 to 2016. A new proxy for financial integration is introduced: the volume of bank-specific Google queries obtained from Google Trends. It is assumed that Google Trends will show higher volumes of queries in a specific region if a bank is more integrated in that region. This offers a new way to proxy financial integration, which is furthermore compatible with the growing importance of digitalisation that has changed the banking industry as well. Even though this thesis does not provide evidence for the relationship between financial integration as proxied by bank-specific Google queries and performance in the European banking sector, it does bring a new way of using web search data in empirical research to light. As Google Trends can still be explored in numerous ways, suggestions for future research imply using web search data in new manners and different fields of economic research.

Table of Contents

I Introduction

Financial integration has been a topic of considerable interest in the past few years. Its potential benefits are broadly discussed in academic literature. To begin with, financial integration is said to improve cross-border holdings of financial assets. In addition, financial integration could improve the degree of consumption risk sharing. This will offer countries more possibilities in smoothing their consumption inter-temporally in response to country-specific income uncertainties. (Kose, Prasad & Terrones, 2003). At last, the benefits for economic growth are mentioned as well (Obstfeld, 1994; Acemoglu and Zilibotti, 1997). Empirical research has indeed shown positive implications of financial integration (Walkner & Raes, 2005; Demirgüç-Kunt, Levine, & Min, 1998; Claessens, Demirgüç-Kunt, & Huizinga, 2001). Nonetheless, debate still exists on the extent to which a positive influence of financial integration exists. For instance, Edwards (2001) finds that high institutional quality is a requirement for countries to benefit from financial integration. Likewise, Blomstrom, Lipsey, and Zejan (1994) conclude that only for rich countries one can observe a positive relationship between financial integration and economic growth. The conflicting outcome of financial integration has been further emphasised, especially after the last financial crisis. José Manuel González-Páramo (2010), member of the Executive Board of the European Central Bank, stated that "during the last financial crisis and its aftermath, financial integration may have contributed to undermine market efficiency, fuel systemic risk and exacerbate the cross-border transmission of financial shocks". In empirical research, financial integration has often been measured either using data on cross-border capital flows or indexes that measure cross-border flow restrictions (Edison, Levine, Ricci & Sløk, 2002).

In this thesis, a new proxy for financial integration will be introduced: the volume of bank-specific Google queries. These volumes are obtained from Google Trends. This is used to analyse the relationship between financial integration and performance in the European banking sector. Google Trends offers insights in the volume of queries that Google users have entered in the search engine. Monthly data can be obtained, starting from January 1, 2004. Google Trends only keeps track of queries from which volumes have reached a certain threshold. Furthermore, relevant variations of the queries are also considered in the calculations. In other words, Google Trends uses a broad match. Data on these query volumes can be found per geographic region and for a specific chosen period. (Google Trends, 2017). For instance, one can obtain information on the volume of the query 'BNP Paribas' in Germany

in the first month of 2012. This shows the 'popularity' of the French bank in Germany as measured by Google queries. To introduce this as a new proxy for financial integration, certain assumptions have to be made. First of all, the assumption is made that these searches give an indication of 'interest' with regard to foreign banks. Secondly, it is also assumed that a higher degree of financial integration can be associated with more 'interest' in foreign banks and thus higher volumes of banks-specific Google queries. If for instance BNP Paribas is more integrated in the German banking sector than the Italian banking sector, German consumers can more easily use the services of the BNP Paribas bank. It is therefore more likely that German consumer will search for 'BNP Paribas' in Google more often than Italian consumers. Therefore, Google Trends will show higher volumes of queries in a specific region if a bank is more integrated in that region. This offers a new way to proxy financial integration, which is furthermore compatible with the growing importance of digitalisation that has changed the banking industry as well.

Technological development in the banking sector has made banking in the 21th century a lot more convenient. Full online banks have been introduced as well. For instance, Dutch online bank Knab provides its customers services which are fully online-based. Almost all banks now offer their customers an online platform where they can easily manage their finances and obtain information on various topics. These developments have made it as easy to access information on foreign banks as domestic banks. It would be interesting to consider these developments when analysing the relationship between financial integration and European bank performance. Therefore, this thesis introduces a proxy for financial integration that takes this into account. The following research question will be assessed:

Does financial integration as proxied by bank-specific Google queries influence European bank performance in the period 2007 to 2016?

This thesis aims to contribute to existing literature, by introducing Google queries as a new proxy for financial integration. The remainder of this thesis will first discuss the theoretical framework for the analyses in section II. In this theoretical framework, financial integration will be broadly defined. Section III will cover the data and methodology for the analyses. In this section, the use of Google queries as a new proxy for financial integration will be discussed in more detail. Its positive and negative implications will be focused on. Results of the analyses

will be discussed in section IV. At last, this thesis will end with a discussion of the results and concluding remarks in section V.

II Theoretical Framework

1. Defining financial integration

To start this research a suitable definition for financial integration should be chosen first. Benoit Coeure, member of the executive board of the European Central Bank, describes financial integration in the Eurosystem as "a situation whereby there are no frictions that discriminate between economic agents in their access to – and their investment of – capital, particularly on the basis of their location. This means that financial integration is achieved when there is equal market access, de facto and de jure" (ECB, 2013). This definition of financial integration will be used for this research as it is compatible with the proxy of financial integration that is introduced. Google is accessible to all kinds of agents and for different purposes. It is unclear whether one uses the engine for personal purposes or firm-specific reasons. The definition includes 'economic agents' and therefore suits the scope of this research.

2. The use of web search data in empirical research

Web search data has already been used in empirical studies. Even though it has been used in different fields of studies, this section will focus on the use of web search data in economic research. In the field of economics, it has been often used for forecasting. Choi and Varian (2012) explain how search engine data from Google Insights for Search can be used to forecast automobile demand, initial claims for unemployment benefits and travel destinations. Google Insights for Search was launched in 2008 as an extension of Google Trends, specifically for marketing purposes. In certain areas, it could provide more detailed information on web search data (Google Trends, 2017). Choi and Varian (2012) concluded that "simple seasonal AR models that include relevant Google Trends variables tend to outperform models that exclude these predictors by 5 per cent to 20 per cent". With this outcome, they tried to stimulate the use of web search data in economic research. D'Amuri and Marcucci (2010) found similar results for their forecast of the US unemployment rate when using a Google-job search index. Using

this index as an explanatory variable increased the explanatory power of their forecast. Ettredge, Gerdes and Karuga (2005) used web search data to analyse the unemployment rate in the US as well. Their research showed a positive relationship between employment-specific queries and unemployment levels. Likewise, Askitas and Zimmerman (2010) found a strong correlation between employment-specific queries and unemployment levels in Germany. In addition, their forecast of the unemployment rate based on web search data seemed to move quite well together with the actual unemployment rate. Guzman (2011) used a similar forecast for the inflation rate. Overall, empirical evidence has shown the potential of web search data in economic research.

3. Empirical evidence on the relationship between financial integration and competition in the European banking sector

Previous empirical research on financial integration in the European banking sector has mainly focused on the effect it has on competition. Molyneux, Lloyd-Williams and Thorton (1994) analysed competitive conditions in the European banking sector in the period 1986 to 1989. Their result showed a lack of financial integration. Therefore, they emphasised the importance of introducing the Second Banking Directive. Under this legislation, "all credit institutions authorized in an EU country would be able to establish branches or supply cross-border financial services in the other countries of the EU without further authorization, provided that the bank was authorized to provide such services in the home state" (Dermine, 2003). Bikker and Groenveld (1998) improved the results of Molyneux et al. (1994). They claim that gradual changes in competition in the European banking sector were not considered in the analyses. Therefore, Molyneux et al. (1994) did not provide stable results. Considering the changes that occurred in the period 1989 to 1996, Bikker and Groenveld (1998) provided new insights. Their result showed that the Second Banking Directive did not increase competition in the European banking sector. However, their analyses also showed that the level of competition was already high prior to the creation of the Single Market. Likewise, Casu and Molyneux (2003) did not provide evidence for an increase in efficiency in the European banking sector after the implementation of the Second Banking Directive. Low efficiency levels were mainly influenced by country-specific factors. In contradiction to the results above, Walkner and Raes (2005) showed that integration and consolidation in the European Union can improve the banking sector through higher levels of efficiency and consumer welfare.

4. Empirical evidence on the relationship between financial integration and bank performance

Many banks worldwide have expanded internationally in the past few decades by establishing branches in foreign countries. Less restrictions in financial markets enhance the internationalization in the banking industry (Claessens et al., 2001). The implications of financial integration in banking sectors have received much attention in academic literature. Levine (1996) states that foreign bank entry could improve efficiency in a banking sector through increased competition. The increasing competition could lead to a decrease in profits and overhead expenses, which is often used to measure the level of efficiency. Higher values of profits and overhead expenses thereby show a lower level of efficiency (Demirgüç-Kunt et al., 1998). Claessens et al. (2001) showed that foreign entry increases the efficiency in the domestic banking sector. Foreign banks import their knowledge and expertise, which is beneficial for the domestic banks. Moreover, foreign entry increases competition and thereby forces more efficient conduct in the overall banking sector. They provided empirical evidence for decreasing profits. The dependent variables 'net non-interest income and total assets ratio', 'before tax income and total assets ratio', 'overhead and total assets ratio' and 'loan loss provisions and total assets ratio' were analysed. The results showed a negative relationship between foreign bank presence and all dependent variables. Likewise, Demirgüç-Kunt et al. (1998) also found empirical evidence for the negative relationship between foreign bank entry and profits.

Walkner and Raes (2005) showed that financial integration in Europe could lead to higher levels of efficiency in the banking sector. In addition, it is explained how higher levels of efficiency imply lower profits as shown by various empirical studies. In this research, bank performance will only be measured by income. Performance could also consider efficiency. However, this is not included in this research. A more detailed explanation will follow in the 'Data' section of this thesis. Given these statements, the hypothesis for this thesis is formulated as follows:

H0: There is no relationship between financial integration as proxied by bank-specific Google queries and bank performance in Europe.

Ha: There is a negative relationship between financial integration as proxied by bank-specific Google queries and bank performance in Europe.

As a new proxy for financial integration is introduced, it could also be the case that a positive relationship will be found instead. It is not clear whether this new proxy will have the same implications on profits or income as shown in previous empirical research. As it is assumed that bank-specific Google queries give an indication of 'interest' with regard to these banks, one could argue that a higher degree of 'interest' will lead to higher income.

III Data & Methodology

1. Data

The Dataset

The dataset for this research includes observations on the performance of 31 European banks. These banks belong to the 50 largest banks in Europe in 2016 as measured by their total assets (Statista, 2016). Not all banks in this list are publicly registered. Therefore, report data for some banks is not as easily accessible. For this reason, the dataset only includes banks that are publicly registered. Banks of which the data did not provide extensive coverage for the period 2007 to 2016 were also excluded from the dataset. The list of the 31 banks that were included in this research can be found in Appendix 1. Furthermore, the dataset includes observations on the volume of bank-specific Google queries, which forms the independent variable in this research. The real gross domestic product, interest rate and the inflation rate of the countries in which the banks are located are included as control variables. A dummy variable to account for periods of recession is also included as a control variable. At last, Google queries from the country in which a bank is located is also added as a control variable. The dataset includes quarterly observations for the period 2007 to 2016 for all variables that are used. Even though Google Trends offers data starting from 2004, data for the period 2004 to 2006 could not be used as a lot of banks showed many missing data points for this period. The period 2007 to 2016 was chosen to include the data from as many banks as possible and with extensive coverage. In total, the dataset includes 8680 observations.

Dependent variable

Bank performance is measured by taking the ratio of 'before tax income' relative to 'total assets', which forms the dependent variable in this research. Claessens et al. (2001) also used this ratio as one of the dependent variables in their analyses of the relationship between financial integration and bank performance. Data has been obtained from Compustat Global,

which offers its data through Wharton Research Data Services. From the 1240 observations that had to be obtained, 20 data points were missing for 'total assets'. In order to create a balanced data set, averages were taken to fill in the blanks. It should be noted that for all missing values, the values for the previous quarter and up following quarter were known. Based on the assumption that the amount of 'total assets' does not change drastically per quarter, averages were taken.

Independent variable

The independent variable in this research is the volume of bank-specific Google queries, introduced as a proxy for financial integration. The volume of Google queries for the 31 European banks is given for 16 different European countries. A list of the countries can be found in Appendix 2. Data for this variable has been manually obtained from Google Trends for each bank, per country. As explained, Google Trends uses a broad match. This implies that the query 'BNP' is also included in the information that Google Trends provides on the query 'BNP Paribas'. Google Trends defines the data as follows: "Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. Likewise, a score of 0 means the term was less than 1% as popular as the peak." (Google Trends, 2017).

Values for the dependent variable 'before tax income' and 'total assets' ratio show an aggregated result from all countries in which a bank is active. In contrast, data from Google Trends has been obtained for each bank, per country. In other words, this data is not aggregated. In order to obtain an aggregated form of the data, the gross domestic product is used to weigh the independent variable. Table 1 shows an illustration of how these weighted averages are obtained. BNP Paribas is taken as an example. Data on the volume of Google queries for 'BNP Paribas' coming from 15 different countries have been obtained for the first quarter of 2007. The country in which the bank is located is excluded from this list. In table 1, this is only illustrated for Austria, Belgium and Denmark. The values are multiplied by the ratio of the country's gross domestic product relative to the total gross domestic products of all countries included. Thereafter, a summation of all the 15 values of 'interest' is taken. The independent variable shows a weighted average of all 'interest' coming from 15 European countries in the specific bank and for a specific quarter. The process is repeated for the 30 other European banks for each quarter in the period 2007 to 2016. The independent variable will be denoted by 'Foreign Interest'.

Table 1: Example from the dataset to illustrate how the independent variable 'Foreign Interest' is obtained.

			Volume of Google queries for 'BNP Paribas'	Country's GDP / summation of all GDP's	Interest
Q1 2007	BNP Paribas	Austria	14,33	0,028	0,40124
		Belgium	3,33	0,036	0,11988
		Denmark	31,33	0,024	0,75192

An advantage of using bank-specific Google queries is that it could offer interesting insights as a new proxy for financial integration. In addition, Google Trends provides a quantitative measure that does not discriminate between the users of the search engine. However, this could also be a downside of using web search data. Google Trends keeps track of all queries, but cannot provide information on its users and their reasons for looking up certain key words. It is arguable whether this could negatively affect the accuracy of the analyses. Nevertheless, empirical evidence does stimulate the use of web search data in economic research.

Control variables

Following the research by Claessens et al. (2001) the gross domestic product, interest rate and inflation rate of the countries in which the banks are located will be included as control variables. The gross domestic product concerns the real gross domestic product that is seasonally adjusted. The interest rate concerns the interest rate on government bonds. The gross domestic product, interest rate and inflation rate are all obtained from the database of the International Monetary Fund. Furthermore, the dataset will also include a dummy variable to account for periods of recession as explained. This dummy variable is publicly available and offered by the Research Division of the Federal Reserve Bank of St. Louis. The chosen dummy variable considers recession indicators for OECD Europe. It takes a value of 1 in case of a recession and 0 if not. At last, the control variable 'Domestic Interest' shows the volume of bank-specific Google queries from the country in which the bank is located.

Table 2 provides a summary of what is expected regarding the sign of the coefficients of all variables. As hypothesized, it is expected that there will be a negative relationship between financial integration as proxied by bank-specific Google queries and bank performance in

Europe. The gross domestic product, interest rate and inflation rate are expected to show positive coefficients in the regression analyses following the research by Claessens et al. (2001). It should be noted that Claessens et al. (2001) use first differences for all the variables. As this will not be the case in this research, it is possible that the actual outcome could differ from what is expected. A further explanation on the use of first differences is provided in the following section. Furthermore, it is expected that the coefficient of 'Domestic Interest' will be positive as it is more likely that the 'before tax income' and 'total assets' ratio will increase if people in the domestic country show more interest in the bank. In periods of recession, a lower 'before tax income' and 'total assets' ratio is expected. Therefore, the table shows a negative sign for the coefficient of the variable 'Recession'.

Table 2: Expected signs for all variables

	Before Tax Income / Total Assets
Foreign Interest	-
Domestic Interest	+
GDP	+
Interest Rate	+
Inflation Rate	+
Recession	-

Descriptive statistics

Descriptive statistics for the chosen variables in the dataset are summarised as follows:

Table 1: Descriptive Statistics

Variable	No. of Observations	Mean	Standard Deviation	Min	Max
Before Tax Income / Total Assets	1240	1.073661	2.310221	-20.50628	16.75788
Foreign Interest	1240	39.15082	16.43739	0	76.38204
Domestic Interest	1240	54.30618	25.34134	0	98.66667
GDP	1240	310.8311	203.0301	58.56106	713.2414
Interest Rate	1240	2.774835	1.468972	-0.52	6.613333
Inflation Rate	1240	1.543468	1.394277	-1.4	5.6
Recession	1240	0.475	0.499576	0	1

2. Methodology

This research aims to analyse if there is relationship between financial integration and bank performance in Europe, where financial integration is proxied by bank-specific Google queries. The data that will be used for this research concerns panel data. Panel data consists of observations on the same entities over two or more periods in time. The dataset includes observations on the ratio of 'before tax income' relative to 'total assets' of 31 different European banks for the period 2007 to 2016. Therefore, the entity in this panel dataset is the European bank.

The regression model will take into account fixed effects. This is tested using the Hausman test, which shows whether there is a significant difference between the fixed and random effects estimators. In the dataset that is used for this research, bank-specific fixed effects will control for omitted variables that vary across banks but remain constant over time. In other words, the fixed effects model will control for unobserved bank level heterogeneity. The fixed effects model does not consider time fixed effects, as the aim of this research is to analyse the relationship between financial integration and European bank performance over time.

The entity fixed effects regression model is also a first difference regression model. As described in the 'Data' section of this thesis, the independent variable 'Foreign Interest' is

defined as follows: “Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. Likewise, a score of 0 means the term was less than 1% as popular as the peak.” (Google Trends, 2017). Therefore, an interest value of 64 in bank A, does not have to be equal to an interest value of 64 in bank B. All number are relative to the highest peak. Taking first differences of the variable in the regression can solve this problem and enable the comparison of interest values over entities and over time. It is possible to compare the change in ‘interest’ in bank A relative to the previous period with the change in ‘interest’ in bank B relative to the previous period. The same is also valid for the control variable ‘Domestic Interest’. Given all statements mentioned above, the formal notation of the regression model is as follows:

$$BeforeTaxIncome/TotalAssets_{it} = ß_0 + ß_1\ \Delta ForeignInterest_{it} + ß_2\ \Delta DomesticInterest_{it} + ß_3\ GDP_{it} + ß_4\ InterestRate_{it} + ß_5\ InflationRate_{it} + ß_6\ Recession_{it} + \alpha_{it} + u_{it}$$

i	= 1,…, 31, denoting the 31 European banks that are included.
t	= 1,…, 40, time period, denoting quarterly observations in the period 2007 to 2016.
$\Delta ForeignInterest_{it}$	$= ForeignInterest_{it} - ForeignInterest_{it-1}$
$\Delta DomesticInterest_{it}$	$= DomesticInterest_{it} - DomesticInterest_{it-1}$
α_{it}	= Bank fixed effects
u_{it}	= Error term

The model will be tested for the presence of stationarity using the Harriz-Tzavalis unit-root test. The Harris-Tzavalis test is chosen, because the panel dataset includes data for a relative small time period and many panels (Harris & Tzavalis, 1999). In addition, it is important that perfect multicollinearity is not present. This would imply an interrelationship between two or more independent variables. In other words, a correlation between independent variables can be found. If multicollinearity is present, it not possible to identify which variables have an influence on the dependent variable. Multicollinearity will be tested using the Variance Inflation Factor (VIF). If the Variance Inflation Factor shows a value lower than 2.5 it will be concluded that multicollinearity is not present (Kutner, Neter, & Nachtsheim, 1996). At last, standard robust errors are included to account for heteroscedasticity (Stock & Watson, 2015).

IV Results

Fixed effects regression model 1 in table 3 shows first differences for all variables that are included. The results of the regression show coefficients for the control variables that are inconsistent with the expectations as described in the 'Data' section of this thesis. None of the variables show significance at the 90%-, 95%- or 99% confidence level. The F-statistic of 0.0317 does shows that all coefficients in the model are different than zero. With a rho of 0.0024175, 0.2% of the variance is due to differences across panels. Taking the first differences of all variables has been further analysed in model 2 and 3 to see if the fixed effects regression model could be improved. Model 2 shows a fixed effects regression in which the first difference of the dependent variable is not taken. The results show coefficients that are consistent with the expectations, except for the dummy variable 'Recession'. The adjusted R-squared and rho have increased in comparison to model 1. The F-statistic shows that all coefficients in the model are still different than zero. In a further analysis, the first differences of the control variables were also removed. Model 3 shows the results of a fixed effects regression model that only includes first differences of the independent variable and the control variable 'Domestic Interest'. The first differences of these variables are taken in order to interpret the coefficients as described in the 'Methodology' section of this thesis. Model 3 shows coefficients that are consistent with the expectations, except for the variable 'Interest Rate'. Relative to model 1 and 2, models 3 shows the highest adjusted R-squared and rho. Furthermore, all coefficients in the model are still different than zero considering the F-statistic. Based on these findings, further analyses have only included first differences of the independent variable and the control variable 'Domestic Interest'. Fixed effects regression model 3 forms the basis for further analyses. It should be noted that these comparisons were made using fixed effects regression models even though a Hausman test was not consulted yet. Fixed effects were still included as these are also included in the remainder of the analyses where a Hausman test will be used. A similar comparison of models that do not include fixed effects can be viewed in Appendix 3. Appendix 3 does not show strikingly different results as compared to the table 3. Model 3 is still used as a base model for the remainder of the analyses. Therefore, the results in Appendix 3 will not be described extensively in this section.

Table 3: Comparison of fixed effects regression models when including and excluding first differences for variables in the analyses.

Dependent variable: Δ (Before Tax Income / Total Assets)		Dependent variable: Before Tax Income / Total Assets		Dependent variable: Before Tax Income / Total Assets	
	(1)		(2)		(3)
Δ Foreign Interest	-0.0070613	Δ Foreign Interest	0.0075074	Δ Foreign Interest	0.0123554
	(0.0134166)		(0.0085705)		(0.0087314)
Δ Domestic Interest	0.010253	Δ Domestic Interest	0.0020464	Δ Domestic Interest	0.0004576
	(0.0071348)		(0.0056164)		(0.0054672)
Δ GDP	-0.0013616	Δ GDP	0.0261633*	GDP	0.0106125
	(0.0098398)		(0.0146489)		(0.0093634)
Δ Interest Rate	-0.2090949	Δ Interest Rate	0.3560931	Interest Rate	-0.0312436
	(0.3666775)		(0.2983592)		(0.0984021)
Δ Inflation Rate	-0.1738806	Δ Inflation Rate	0.2581728**	Inflation Rate	0.0501909
	(0.0716954)		(0.1217665)		(0.0715159)
Δ Recession	0.0304926	Δ Recession	0.1946977	Recession	-0.554765***
	(0.138201)		(0.1317007)		(0.1540575)
Constant	-0.1076592	Constant	1.036665***	Constant	-2.000939
	(0.0340162)		(0.0302516)		(3.075719)
No. of observations	1240	No. of observations	1240	No. of observations	1240
No. of banks	31	No. of banks	31	No. of banks	31
Adjusted R-squared	0.0085	Adjusted R-squared	0.0198	Adjusted R-squared	0.1110
Prob. (F-statistic)	0.0317	Prob. (F-statistic)	0.0062	Prob. (F-statistic)	0.0236
Rho	0.00247175	Rho	0.20693019	Rho	0.63874957
Bank fixed effects	yes	Bank fixed effects	yes	Bank fixed effects	yes

Legend:

Δ = change relative to the previous period, t-1.

Standard errors are reported in parentheses. *, **, *** indicating significance at the 90%, 95%, and 99% level, respectively

Table 4 shows the results of model 3 next to the results of regression model 4. Model 4 shows the results of a regression in which fixed effects are not included. Model 3 in table 4 is exact the same as model 3 in table 3. It is presented again to easily compare the regression model with and without fixed effects. Model 4 shows a positive relationship between the change in foreign interest relative to the previous period and the 'before tax income' and 'total assets' ratio. However, the coefficient for the independent variable is insignificant. Coefficients for the control variables 'Interest Rate', 'Inflation Rate' and 'Recession' are consistent with the

expectations as described in the 'Data' section of this thesis. However, a significant relationship can only be observed for control variable 'Recession'. The result shows a negative and significant relationship between GDP and the dependent variable, which is inconsistent with the expectations. The coefficient for the control variable 'Domestic Interest' does show a positive sign as expected. However, the coefficient is insignificant. Furthermore, the coefficient for the constant is negative and significant. All coefficients in the model are different than zero as shown by the F-statistic. Fixed effects increase the adjusted R-squared, indicating that model 3 fits the data better than model 4. For this reason, model 3 has been used for further analyses.

Fixed effects regression model 3 controls for unobserved bank level heterogeneity that is constant over time. The results show a positive, but insignificant relationship between the change in foreign interest and the 'before tax income' and 'total assets' ratio. All control variables apart from 'Interest Rate' show signs that are consistent with the expectations in this thesis. It should be noted that the control variable 'Recession' is the only variable with a significant coefficient. The F-statistic of 0.0236 shows that all coefficients in the model are different than zero. The rho shows that 64% of the variance is due to differences across panels. Model 3 has been tested for stationarity. As explained, the first differences of the independent variable and the control variable 'Domestic Interest' are taken to enable the comparison of values over entities and over time. This can also take out the unit root if it is present in the time series. The Harris-Tzavalis unit root test for the fixed effects regression model showed that stationarity was present before and after using first differences. Furthermore, table 5 shows the value of the Variance Inflation Factor for the independent variable and control variables. The value of the average Variance Inflation Factor is also given. The outcome shows that perfect multicollinearity is not present. At last, a modified Wald test showed the presence of heteroscedasticity in the dataset. Therefore, robust standard errors have been applied.

With a positive and insignificant relationship between the change in foreign interest and the 'before tax income' and 'total assets' ratio, the null hypothesis is not rejected in favour of the alternative. It should be noted that this is also the case when considering model 4.

Table 4: Results with and without bank fixed effects.

Dependent variable: Before Tax Income / Total Assets		
	(4)	(3)
Δ Foreign Interest	0.0145776	0.0123554
	(0.0106674)	(0.0087314)
Δ Domestic Interest	-0.0031736	0.0004576
	(0.0088824)	(0.0054672)
GDP	-0.0016274***	0.0106125
	(0.0002297)	(0.0093634)
Interest Rate	0.0250316	-0.0312436
	(0.0637771)	(0.0984021)
Inflation Rate	0.0101146	0.0501909
	(0.0646748)	(0.0715159)
Recession	-0.5195503***	-0.554765***
	(0.1288378)	(0.1540575)
Constant	1.703335***	-2.000939
	(0.1628554)	(3.075719)
No. of observations	1240	1240
No. of banks	31	31
Adjusted R-squared	0.0350	0.1110
Prob. (F-statistic)	0.0000	0.0236
Rho	-	0.63874957
Bank fixed effects	no	yes
Hausman	-	0.0537

Legend:

Δ = change relative to the previous period, t-1.

Standard errors are reported in parentheses. *, **, *** indicating significance at the 90%, 95%, and 99% level, respectively

Table 5: Variance Inflation Factor (VIF) for fixed effects regression model 3.

Variable	Variance Inflation Factor (VIF)
Δ Foreign Interest	1.07
Δ Domestic Interest	1.07
GDP	1.01
Interest Rate	1.46
Inflation Rate	1.55
Recession	1.10
Mean Variance Inflation Factor (VIF)	1.21

Model 3 has also been analysed to see whether a fixed effects regression model is indeed more suitable than a random effects regression model. This analysis has been conducted with the Hausman test. Results of the Hausman test showed insignificance of 0.0537 as presented in table 4. Therefore, further analyses have been conducted to see if model 3 could be improved. Table 6 shows the results of a fixed effects regression model in which 'GDP' and 'Recession' are the only control variables. Results of the Hausman test show a significance of 0.0236, indicating that a fixed effects model is indeed the most suitable model. Furthermore, stationarity is present and perfect multicollinearity is not found as already presented in table 5. Robust standard errors have been used to account for heteroscedasticity. A positive, but insignificant relationship between the change in foreign interest and the 'before tax income' and 'total assets' ratio can be observed. Signs of coefficients for the variables 'GDP' and 'Recession' are as expected. Similar to model 3, 'Recession' is the only variable with a significant coefficient. The F-statistic of 0.0038 shows that all coefficients in the model are different than zero. With a rho of 0.65101245, 65% of the variance is due to differences across panels. It has increased relative to the rho in model 3. On the contrary, the adjusted R-squared has decreased.

With a positive and insignificant relationship between the change in foreign interest and the 'before tax income' and 'total assets' ratio, the null hypothesis is still not rejected in favour of the alternative.

Table 6: Results fixed effects regression model including 'GDP' and 'Recession as control variables.

Dependent variable: Before Tax Income / Total Assets	
	(5)
Δ Foreign Interest	0.0123934
	(0.0091533)
GDP	0.0109959
	(0.0074677)
Recession	-0.5256805***
	(0.1412585)
Constant	-2.137846
	(2.303855)
No. of observations	1240
No. of banks	31
Adjusted R-squared	0.1125
Prob. (F-statistic)	0.0038
Rho	0.65101245
Bank fixed effects	yes
Hausman	0.0236

Legend:

Δ = change relative to the previous period, t-1.

Standard errors are reported in parentheses. *, **, *** indicating significance at the 90%, 95%, and 99% level, respectively

V Conclusion

This research has been conducted to formulate an answer to the following research question: 'Does financial integration as proxied by bank-specific Google queries influence European bank performance in the period 2007 to 2016?' In answering this question, one hypothesis has been phrased to analyse the relationship between financial integration as proxied by bank-specific Google queries and bank performance in Europe. The results of fixed effects regression model 3 have shown that the hypothesis cannot be rejected in favour of the alternative. The hypothesis states that there is a negative relationship between financial integration as proxied by bank-specific Google queries and bank performance in Europe. The same conclusion holds when considering both models 4 and 5. The findings in this thesis do not provide evidence for the relationship between financial integration as proxied by Google queries and performance in the European banking sector in the period 2007 to 2016. Limitations in this research could have influenced these findings.

First of all, limited availability of data has formed one of the main limitations in this research. The dataset included quarterly observations for the period 2007 to 2016 for all variables that were used. Even though Google Trends offers data starting from 2004, data for the period 2004 to 2006 could not be used as a lot of banks showed many missing data points for this period. Furthermore, the dataset only included banks that are publicly registered. Banks of which the data did not provide extensive coverage for the period 2007 to 2016 were also excluded from the dataset. The dataset eventually included observations on the performance of 31 European banks. Suggestions for future research on this topic imply using a larger scope. If more data will be available, the analyses could be conducted over longer period of time. In addition, more European banks could be included in the dataset as well.

The results give an indication that the volume of bank-specific Google queries might not have been a good proxy for financial integration. Even though empirical evidence has shown the potential of Google Trends in economic research, it could be the case that it is not as suitable in the analyses of financial integration and bank performance. Furthermore, data in this research has been aggregated using the gross domestic product. The independent variable showed a weighted average of all 'foreign interest' coming from 15 European countries in a specific bank and for a specific quarter. One could argue whether this has been an accurate way to aggregate the data and if any other methods would have been more suitable. Moreover, future research could include the observations of bank-specific Google queries for more countries. Furthermore, this thesis only considered one variable to assess European bank performance: the ratio of 'before tax income' relative to 'total assets'. Suggestions for future research also include the use of more variables to analyse bank performance.

Even though the results in this thesis might not have shown strikingly, new implications in the field of financial integration and banking, this thesis can stimulate the use of Google Trends in new manners and in different fields of economic research. The possibility exists that future research can find new insights when the scope is broadened and more data will be available. Furthermore, Google Trends in itself, rather than as a proxy for financial integration, could still provide interesting insights in banking sector analyses. As the use of web search data is relatively new in empirical research, it can still be explored in numerous ways. With the introduction of bank-specific Google queries as a new proxy for financial integration, this thesis did not only aim to contribute to existing literature, but also offered new insights in how Google Trends could be used in economic research.

References

Acemoglu, D., & Zilibotti, F. (1997). Was Prometheus Unbound by chance? Risk Diversification, and Growth. *Journal of Political Economy, 105*(4), 709-51.

Askitas, N, & Zimmermann, K. F. (2010). Google Econometrics and Unemployment Forecasting. *Applied Economics Quarterly, 55*(2), 107-120.

Bikker, J. A., & Groenveld, J. M. (1998). *Competition and Concentration in the EU Banking Industry.* Retrieved on May 1, 2017, from https://www.dnb.nl/binaries/ot008_tcm46-146043.pdf

Blomstrom, M., Lipsey, R., & Zejan, M. (1994). What Explains the Growth of Developing Countries? In J Baumol, R.R. Nelson, & E.N. Wolf (Eds.), *Convergence of Productivity* (pp.243-262). New York, NY: Oxford University Press, Inc.

Casu, B., & Molyneux, P. (2003). A comparative study of efficiency in European banking. *Applied Economics, 35*(17), 1865-1876.

Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *The Economic Record*, 88, 2-9.

Claessens, S., Demirgüç-Kunt, A., & Huizinga, H. (2001). How does foreign entry affect domestic banking markets? *Journal of Banking & Finance, 25*(5), 891-911.

D'Amuri, F, & Marcucci, J. (2010). *Google it! Forecasting the US Unemployment Rate with a Google Job Search Index.* Retrieved on July 10, from https://www.econstor.eu/bitstream/10419/43536/1/640335640.pdf

Demirgüç-Kunt, A., Levine, R., & Min, H. G. (1998). *Opening to foreign banks: Issues of stability, efficiency, and growth.* Retrieved on May 1, 2017, from https://www.researchgate.net/profile/Hong-ghi_Min/publication/305386836_Opening_to_foreign_banks_Issues_of_stability_efficiency_and_growth/links/5847e35b08aeda696825a75c/Opening-to-foreign-banks-Issues-of-stability-efficiency-and-growth.pdf

Dermine, J. (2003). Banking in Europe: Past, Present, and Future. In V. Gaspar, P. Hartmann, & O. Sleijpen (Eds.), *The Transformation of the European Financial System* (pp.31-118). Frankfurt am Main, Germany: European Central Bank.

FRED Economic Research Federal Reserve Bank of St. Louis. (2017). *Recession Indicators Series.* Retrieved on May 1, 2017, from https://fred.stlouisfed.org/release?rid=242

Edison, H. J., Levine, R., Ricci, L., & Sløk, T. (2002). International financial integration and economic growth. *Journal of international money and finance, 21*(6), 749-776.

Edwards, S. (2001). *Capital Mobility and Economic Performance: Are Emerging Economies Different?* Retrieved on May 1, 2017, from https://core.ac.uk/download/pdf/6606196.pdf

Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using Web-based Search Data to Predict Macro-economic Statistics. *Communications of the ACM, 48*(11), 87–92.

European Central Bank. (2013). *The way back to financial integration*. Retrieved on May 1, 2017, from https://www.ecb.europa.eu/press/key/date/2013/html/sp130312.en.html

European Central Bank. (2010). *Globalisation, international financial integration and the financial crisis: The future of European and international financial market regulation and supervision.* Retrieved on May 1, 2017, from https://www.ecb.europa.eu/press/key/date/2010/html/sp100219.en.html

Google Trends. (2017). *Explore*. Retrieved on May 1, 2017, from https://trends.google.com/trends/explore

Guzman, G. (2011). Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations. *The Journal of Economic and Social Measurement, 36*(3), 119-167.

Harris, R. D. F., & E. Tzavalis. (1999). Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics, 91*, 201–226.

International Monetary Fund. (2017). *IMF DATA access to macroeconomic & financial data. Retrieved* on May 1, 2017, from http://data.imf.org/?sk=388DFA60-1D26-4ADE-B505-A05A558D9A42

Kose, M. A., Prasad, E. S., & Terrones, M. E. (2003). *Financial integration and macroeconomic volatility*. Retrieved on May 1, 2017, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=393420

Kutner, M. H., Neter, J., Nachtsheim, C. J., & Li, W. (1996). *Applied Linear Statistical Models*. Chicago, IL: Mcgraw-Hill Education.

Levine, R. (1996). Foreign banks, financial development, and economic growth. In E.B. Claude (Ed.), *International Financial Markets: Harmonization versus Competition* (pp.224-254). Washington, D.C., WA: AEI Press.

Molyneux, P., Lloyd-Williams, D. M., & Thorton, J. (1994). Competitive conditions in European banking. *Journal of Banking & Finance, 18*(3), 445-459

Obstfeld, M. (1994). Risk-taking, Global Diversification, and Growth. *American Economic Review, 84*(5), 1310-1329.

Statista. (2017). *Leading banks in Europe as of 2016, by total assets (in billion U.S. dollars).* Retrieved on May 1, 2017, from https://www.statista.com/statistics/383406/leading-europe-banks-by-total-assets/

Stock, J. W., & Watson, M. W. (2015). *Introduction to Econometrics*. Essex, England: Pearson Education Limited.

Walkner, C., & Raes, J. P. (2005). *Integration and consolidation in EU banking-an unfinished business*. Retrieved on May 1, 2017, from http://ec.europa.eu/economy_finance/publications/pages/publication662_en.pdf

Wharton Research Data Services. (2017). Compustat Global. Retrieved on May 1, 2017, from https://wrds-web.wharton.upenn.edu/wrds/

Appendix

1.

List of European Countries	
Austria	Luxembourg
Belgium	The Netherlands
Denmark	Norway
Finland	Portugal
France	Spain
Germany	Sweden
Ireland	Switzerland
Italy	United Kingdom

2.

List of European Banks	
Banco Bilbao Vizcaya Argentaria	ING Group
Banco Sabadell	Intesa Sanpaolo
Barclays	KCB Bank
BNP Paribas	Lloyds Banking
CaixaBank	Nationwide Building Society
Commerzbank AG	Nordea
Crédit Agricole	Raiffeisen Zentralbank Österreich
Credit Suisse Group	Santander Group
Danske Bank	Skandinaviska Enskilda Banken
Deutsche Bank AG	Sociéte Générale
Dexia Group	Standard Chartered
DNB ASA	Swedbank
DZ Bank	The Royal Bank of Scotland
Erste Bank der österreichischen Sparkassen AG	UBS
Handelsbanken	UniCredit
HSBC Holdings PLC	

3.

Comparison of regression models when including and excluding first differences for variables in the analyses (without fixed effects).

Dependent variable: Δ (Before Tax Income / Total Assets)		Dependent variable: Before Tax Income / Total Assets		Dependent variable: Before Tax Income / Total Assets	
	(1)		(2)		(3)
Δ Foreign Interest	-0.0073082	Δ Foreign Interest	0.0115577	Δ Foreign Interest	0.0153236
	(0.0112936)		(0.0113053)		(0.0108176)
Δ Domestic Interest	0.0102216	Δ Domestic Interest	0.0013703	Δ Domestic Interest	-0.011299***
	(0.0101312)		(0.0101418)		(0.0030317)
Δ GDP	-0.0002626	Δ GDP	0.0182634	GDP	-0.001401***
	(0.0142714)		(0.0142862)		(0.0003231)
Δ Interest Rate	-0.2173349	Δ Interest Rate	0.3698283*	Interest Rate	-0.0738946
	(0.2075656)		(0.2077819)		(0.0590734)
Δ Inflation Rate	-0.1724052	Δ Inflation Rate	0.272222**	Inflation Rate	0.0046652
	(0.1123642)		(0.1124813)		(0.0570523)
Δ Recession	0.0309604	Δ Recession	0.1853587	Recession	-0.511622***
	(0.1690235)		(0.1691996)		(0.1346388)
Constant	-0.1091309	Constant	1.044974***	Constant	2.525744***
	(0.0691448)		(0.0692168)		(0.2850235)
No. of observations	1240	No. of observations	1240	No. of observations	1240
No. of banks	31	No. of banks	31	No. of banks	31
Adjusted R-squared	0.0052	Adjusted R-squared	0.0198	Adjusted R-squared	0.0459
Prob. (F-statistic)	0.3952		0.0041		0.0000
Bank fixed effects	no	Bank fixed effects	no	Bank fixed effects	no

Legend:

Δ = change relative to the previous period, t-1.

Standard errors are reported in parentheses. *, **, *** indicating significance at the 90%, 95%, and 99% level, respectively