Which Google keywords influence entrepreneurs? Empirical evidence from Vietnam

Toan Luu Duc Huynh
School of Banking, University of Economics Ho Chi Minh City, Ho Chi Minh City, Vietnam

Received 14 November 2018
Revised 13 June 2019
15 June 2019
16 June 2019
Accepted 17 June 2019

Abstract

Purpose – This paper aims to shed light on an impact of Google keywords on the number of new businesses (and an amount of capital registered) in Vietnam, the Southeast Asian country, after the year of an entrepreneur, 2016.

Design/methodology/approach – This study uses a rich set of quantitative techniques from VAR Granger and threshold regression. The whole sample period covers the data (keywords, number of new businesses, an amount of capital invested to register) from the first week of 2016 to October 2018, which includes 144 observations in total.

Findings – The findings suggest that the relationship between Google does not persist in the long run. There is a short-run shock, might cause a change to the frequency of the other keywords rather than the number of firms (or an amount of capital). However, under the number of firms' threshold, keywords have the both positive and negative impacts on entrepreneurs whereas a higher threshold of capital, keywords show their roles to predict an amount of money for registering firms.

Practical implications – The Vietnamese Government and executives are advised to consider the Google keywords "entrepreneur" (in Vietnamese) and "start-up", which cause a decline in entrepreneurial movements. In addition, the current period is going to inverse from the previous one in terms of the number of firms and an amount of capital. Finally, there are two critical thresholds: 1,602 companies and 35,010m VND for the keywords' influence.

Originality/value – This study contributes empirical evidence of technological change and entrepreneurship and contributes to the existing literature by discussing how this relationship under the threshold.

Keywords Vietnam, Capital, VAR, Entrepreneurs, Threshold, Google search, Granger

Paper type Research paper

1. Introduction

After "*Doi Moi*" (the name given to the economic transformation in Vietnam until 1986), Vietnam changed its policies from the centrally planned economy to the emphasis of

JEL classification – M10, M19, M51

This research is fully funded by University of Economics Ho Chi Minh City (Vietnam). The author thanks Prof. Chang Won Lee and two referees for their valuable comments. Any remaining errors are the author's own responsibility.

Asia Pacific Journal of Innovation and Entrepreneurship
Vol. 13 No. 2, 2019
pp. 214-230
Emerald Publishing Limited
2398-7812
DOI 10.1108/APJIE-11-2018-0063

privatization sectors. The small and medium enterprises (SMEs) in Vietnam contribute to half of the total national GDP. Interestingly, Brandt, Rand *et al.* (2016) indicate that the SMEs and private sectors have been playing the prominent in the growth of Vietnamese economic. Therefore, the Vietnamese Government is going to generate many policies as well as campaigns to foster the nationally entrepreneurial ecosystem. Since 1997, Vietnam has used the Internet as a member of the global network. According to Statista Portal (Statistics and Studies from more than 22,500 Sources), Vietnam currently has 55.19 million users. Thus, it is not difficult for those who want to find the information through the search engines as well as an Internet system in Vietnam. As other countries, Google is one of the Vietnamese favorite tools. Google Trend declares that they record the users' history of finding and searching for statistical purpose. Thus, they can pick how frequency the users want to capture keywords by ranking them in the range from 01 to 1000. The more they search, the higher the indicators are. Before doing business, people have a tendency to find whether there exist potential markets for them to penetrate or not. Hence, we can use Google tools for searching the macroscopic determinants as well as microscopic factors. Can that tool use to predict for entrepreneurial activities in Vietnam? Can the groups of these factors such as high internet penetration, high internet feasibility, and high utilization of Google's search forecast what people react under the information of new business? Our paper is to investigate how Google search influence the entrepreneur on both sides, such as the number of firms and an amount of capital. This is a standard reference for not only academic researchers but also the policymakers when promulgating any policies regarding entrepreneurial movements.

The paper proceeds as follows: Section 2 briefly acknowledges the existing evidence on the subject of the searching engine and entrepreneurial process. Section 3 presents the empirical framework regarding methodology and details on the dataset. Section 4 presents and discusses the empirical findings. Section 5 presents the conclusion and discussion of policy implications.

2. Literature review

It is undeniable that technological change has intervened in almost all areas ranging from medical treatment to economics, which leads to high adaptability by using these widgets. The prediction of technological determinants and entrepreneurship is attracting not only academic researchers but also a businessmen. Google is probably using the most frequent tools for those who apply the searching engine in their lives and work. Most entrepreneurial executives often try to reach those customers, markets, and politics to find their opportunities whether they use their money to commence businesses or not.

The technological change is fostering the current economy by restructuring and repositioning the roles of participants in markets. The Internet is employed as a tool for finding the customers' needs and reaching new customers, which supports to achieve the companies' objectives. There are some studies, which adapts the technological movements to financial models such as Bijl (2016) for forecasting the stock return, Neri *et al.* (2018) to evaluate the stock market activities from Norway or the Gwilym *et al.* (2016) contribution from the Chinese markets. Many scholars can use Google search to predict their socio-economics, such as the unemployment rate under the study of D'Amuri *et al.* (2017). As regards economic policy uncertainty, Baker *et al.* (2016) employ the Google index for predicting employers' and consumers' behaviors on employment dynamics. Not only using Google but also employing Twitter is a typical study from Bollen *et al.* (2011) to investigate for the stock market.

Many previous types of research focus on the personal characteristics or demographic factors, for example, Nga and Shamuganathan (2010) and Scott (2006) or some studies on institutions and entrepreneurship from Acs *et al.* (2008). However, it is still a gap for using Google Search Engine as a tool to forecast the potential of new businesses as well as the capital invested in markets through the decision's making. From another perspective, Scott (2009) opposes that policymakers should not encourage more people to become entrepreneurs, which results in bad public policy. It is illogical for those who use the internet to find the most successful companies rather than their failures in doing business. In addition, Courtois *et al.* (2018) indicate that the Google search engine does not support for solving the social and political information.

Why do people search? Yang *et al.* (2019) referred to the many definitions of searching (Li *et al.*, 2013; Huber, 1991; Kornish and Ulrich, 2011). There is a process that people can learn from failure, find something new, and develop innovation and many different layouts of purposes. However, in the entrepreneurial context, Yang *et al.* (2019) summarized that the searching process in startup attempts to find information to examine their products by proposing them to customers in specific markets. Therefore, search activities can result from many various forms such as the internet, market survey, or interviewing section. Recently, the study of Google search has caught many scholars' attention. Carnes *et al.* (2019) introduced that traditional research organization tends to transform through public entrepreneurship. Then, using the Google search engine is one of their approaches to get further informational questions and possible answers. Yan and Guan (2019) contribute to the existing literature by demonstrating that Google trend is likely to be factors improving entrepreneurial ecosystem cognition in the scope of OECD countries between 2005 and 2014. Noticeably, Askitas and Zimmermann (2009) confirmed that the Google search engine could be used to forecast economic behavior. Hence, this paper attempts to use Google as a proxy to predict the number of businesses in Vietnam through their behaviors. Why is Google? Firstly, Google is still the favorite search engine for the internet with a wide range of users in different countries and regions. Secondly, we based on the assumption from Da *et al.* (2011) that the searchers are limited to attention. Therefore, the Google Trends (Google Search Engine figures) will provide insightful information regarding active internet search of entrepreneurial activities. Thirdly, Davis and Harveston (2000) indicated that most entrepreneurs collected data from the internet, which becomes a popular and common way to obtain what they need. Meanwhile, Xiang and Gretzel (2010) argued that information from search engine reflects heterogeneous information needs. From this logic, this paper attempts to investigate which keywords drive the entrepreneurs in terms of their behaviors by establishing new companies because entrepreneurial information is very important for doing business (Gifford, 1992).

More interestingly, Al and Fazal (2018) also concluded that entrepreneurial orientation could influence the competency and micro-enterprise performance. In particular, the improvement of creativity and innovativeness could enhance micro-enterprise performance as well as inspire poor households to execute entrepreneurial activities. This is also close literature in terms of the ability to use creativity and innovativeness. The entrepreneurs can take their advantage to find the necessary information based on the searching engine platform, which brings them more perspectives that are different in initiating new businesses. Furthermore, Tran and Von Korflesch (2016) confirmed that the supporting perception from the social cognitive career theory is also one of the important elements to cause the intention of establishing new social businesses. Currently, in the internet booming context, the searching engines, as well as other platforms, drive human perception, known as the impacting factor on entrepreneurs' intentions. The study of Lee and Chun (2017)

contributes to the existing literature by building the dynamic ontology through Wikipedia to measure national image. These scholars indicated the role of the internet and relevant online information in terms of driving the human perception regarding the national image. Recently, Ribeiro, and Plonski (2019) used an in-depth interview to show that the entrepreneurship ecosystems can be influenced by the technologies, which includes the searching engine. In brief, there are some several studies investigate the indirect relationship between technological development, particularly searching engine, and entrepreneurial intention. Hence, our paper contributes to the existing literature by adding the Google search engine as a factor explaining the entrepreneurs in a specific context.

In the context of the booming of entrepreneurial movements, people can approach information regarding the 'new business registration' or "start-up' everywhere from communication channels or news. Vietnam has become one of the world's most entrepreneurial nations (ranked #5 in the Approved Index in 2015 from www.approvedindex.co.uk/). Thus, many people are searching for their potential markets before making their decisions on doing business as the study of Kirihata (2018). This research emphasized that not only entrepreneurs but also investors need to have necessary insights before the early-stage. It is too late to find information as well as the procedure in the beginning stage of the entrepreneurship process. Thus, the available information is a channel to transmit their intentions to do business, which might be used as a determinant to predict the business actions in terms of new companies as well as their capital invested. It must be frankly said that overwhelming information is negatively or positively contributed to any economic aspects. Therefore, we decide to employ a rich set of established empirical approaches (included threshold regression) to debate and understand how Vietnamese entrepreneurial markets move after the national campaign.

3. Data and methodology

3.1 Data collection

Our data used is obtained the Google Trends (for the value of searching level) and the gate of Vietnamese Government (specifically Ministry of Planning and Investment) in registering new businesses, starting from the first week of 2016 to the end of October 2018. The main reason for choosing this period for our research is related to political strategy from the Vietnamese Government. At the beginning of the year 2016, the Vietnamese Government officially declares the project called "Startup Nation", which aims to foster the number of new business reaching at 1 million new firms within the country by 2020. After launching this project, the Vietnamese Government's propaganda is to target the multimedia communication for spreading this information, which leads to the frequency of searching in Google as regards entrepreneurship.

3.2 Data processing

Our method for processing the data is based on Bijl *et al.* (2016) and Da *et al.* (2011). This research proves that the average of the previous 52-week data is subtracted from the raw data. Thus, we also perform the standardized by dividing their discrepancy with standard deviation by the following formula (denoted AGSV, Average Google Searching Value Index; GSVI, Google Searching Value Index):

$$\mathrm{AGSVI}_{t}^{B} = \frac{\mathrm{GSVI}_{t} - \frac{1}{52}\sum_{i=1}^{52}\mathrm{GSVI}_{t-1}}{\sigma_{\mathrm{SVI},t}} \quad (1)$$

In which, $AGSVI_t^B$ is Average Google Searching Value Index at week t whereas $GSVI_t$ is Google Searching Value Index at week t and $GSVI_{t-1}$ is Google Searching Value Index at week $(t - 1)$. The main reason why we cannot use the Google Search Value Index for our estimation is the value depending on the period of downloaded data (Neri *et al.*, 2018). In addition, the number of new firms and an amount of capital invested into the Vietnamese market by the percentage of change from the previous week and current week.

3.3 Data analysis

To commence analyzing in the following section, we run a descriptive statistical analysis to get an insight into the features of the data (Table I).

We suggest having further validation such as quantile regression and threshold regression for investigating the insights which keywords drive a change in a number of new business establishment or registered capital (Figure 1).

Interestingly, there is a huge shock in a number of new business establishment and chartered capital at the end of 2016 (the beginning of 2017) over a period from Week 1, 2016 to Week 40, 2018. The main explanation for this phenomenon results from the "January effect" proposed by Thaler (1987), which simulates to the strong motivation for people in the first month of the year. Especially, this phenomenon is intuitively explained by the Asian markets by the chance of Lunar New Year to have a new starting point from a new business. In addition, the movement of a pair keyword between "Entrepreneurship" and "New business" is in-line. Similarly, this is also well-matched with a couple of keyword between "Entrepreneurship' (in Vietnamese) and "Start-up".

3.4 Methodology

We use a rich set of quantitative techniques to estimate whether keywords cause a change in the number of new business by shocks and threshold.

As regards "Granger-causes" and Vector Auto Regression (VAR), we refer to Lopez and Weber (2017), Lütkepohl (2005) and Granger (1969) for time-series estimation.

Based on Hansen (2000)'s study, we propose the approachable models as follows:

$$y_i = \theta'_1 x_i + e_i \text{ for } q_i \leq \gamma \tag{2}$$

$$y_i = \theta'_2 x_i + e_i \text{ for } q_i > \gamma \tag{3}$$

In which, q represents the threshold variable used to separate the sample into one or more regions; x is an m-vector of repressors' and e is the error term. Then, the algorithm for estimation coefficients is minimized the least squares of the following regression with spitted observations and two regions:

Table I. Statistical description for variables

Variables	Explanation	Mean	SD	Min	Max
ENT	Entrepreneurship	−0.0049972	0.9983	−2.129205	4.918857
NB	New business	0.0094991	1.002996	−2.042225	2.765448
ENTVN	Entrepreneurship (Vietnamese)	−0.013445	1.009474	−2.284666	4.164077
SU	Start-up	−0.0199718	1.024912	−2.875941	3.084342
PERNUM	% new business	0.1141855	0.9667872	−0.857466	10.80159
PERCAP	% capital new business	0.2282415	1.306732	−0.9780954	14.07992

$$y_i = \theta'_1 x_i + z_i \delta_1 I(q_i \leq \gamma) + z_i \delta_2 I(q_i > \gamma) + e_i \quad (4)$$

Then, z_i is a vector of exogenous variables with region-specific coefficient vectors δ_1 and δ_2. The estimator for threshold is calculated by:

$$\hat{\gamma} = \arg\min_{\gamma \epsilon T} S_T(\gamma) \quad (5)$$

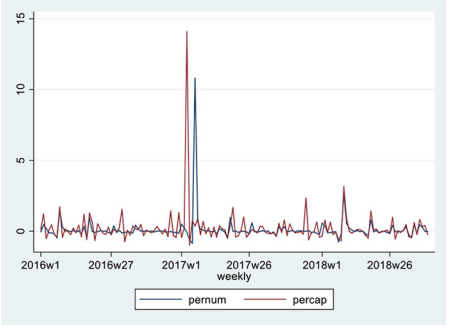
For a sequence of T_1 values in q_i, where $T_1 < T$. It implies that T1 corresponds to the number of observations between the 10th and 90th percentile of q_i. It might be referred to Hansen (2011) for more information on threshold estimation.

4. Results and discussion

4.1 Test of stationary

We use the Dickey and Fuller (1979) test as one of the most commonly used tests for stationarity. The null hypothesis is that the series has a unit root (Table II).

In addition, to ensure that all variables are stationary, we further use the Phillips and Perron (1988) test that a variable has a unit root. The more advantage of Philips–Perron rather than Dickey–Fuller is to integrate Newey and West (1987) standard errors to account for serial correlation (Table III).



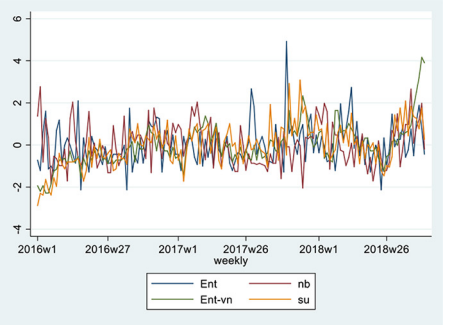


Figure 1. The investigation for percentage in a change of new business establishment and a change in the value of standardized keywords by Google Search Engine

Variables	*t*-statistics	Stationary or not
ENT	−11.895***	Stationary
NB	−9.221***	Stationary
ENTVN	−2.688*	Non-stationary
D.ENTVN	−15.187***	Stationary
SU	−7.284***	Stationary
PERNUM	−12.221***	Stationary
PERCAP	−13.470***	Stationary

Notes: *, ** and *** significant at 10%, 5% and 1% levels, respectively; because of ENTVN being non-stationary in significance level 1%, we use a further process for this variable by taking the first difference of ENTVN, which is considered as one typical way to deal with stochastic trends (unit root); D denotes "first-difference" for original time series

Table II. The test of stationary by Dickey–Fuller (1979)

Variables	*t*-statistics	Stationary or not
ENT	−11.963***	Stationary
NB	−9.448***	Stationary
ENTVN	−2.289	Non-stationary
D.ENTVN	−15.640***	Stationary
SU	−7.442***	Stationary
PERNUM	−12.264***	Stationary
PERCAP	−13.464***	Stationary

Table III.
The test of stationary by Phillips and Perron (1988)

Notes: *, ** and *** significant at 10%, 5% and 1% levels, respectively; because of ENTVN being non-stationary in significance level 1%, we use a further process for this variable by taking the first difference of ENTVN, which is considered as one typical way to deal with stochastic trends (unit root); D denotes "first-difference" for original time series

The alternative specification of unit root tests demonstrates that all variables except ENTVN are stationary at significance level 1 per cent. The first difference of ENTVN (denoted by D.ENTVN) is to reject the null hypothesis for both the Dickeya and Fuller (1979) and the Phillips–Perron (1988) at 99 per cent confidence level. Hence, we use these variables evaluated above for further quantitative analysis.

4.2 Choosing the lag selection by Lütkepohl (2005)

Next, we test of lag-order selection statistics for time series approaches as a pre-estimation. The pre-estimation stage is normally used to choose the lag order for the following appropriate time-series models. We mainly use the study of Lütkepohl (2005) with the represented as the Log-Likelihood (LL), the Likelihood ratio (LR), the Prediction Error (FPE), the Akaike's Information Criterion (AIC), the Schwarz's Bayesian Information Criterion (SBIC) and the Hannan and Quinn Information Criterion (HQIC) (Table IV).

Based on the research of Ivanov and Kilian (2001), the AIC criterion is better for the monthly data for Vector Auto Regression (VAR). Then, they also indicate HQIC is well applicable for quarterly data, which is over 120 observations. Then, SBIC usually works for Vector Error Correction (VEC) model. Therefore, our data will be fitted with a lag of three (3) for further implications of VAR and should be used without lag for VEC model after another test.

4.3 Test of Johansen co-integration

After a choice of the lag selection by Lütkepohl (2005), our estimate the cointegrating rank of a VECM is the following step. We refer to the rich sets of quantitative techniques from

Lag	LL	LR	df	*p*	FPE	AIC	HQIC	SBIC
0	−1,108.58				0.371547	16.0372	16.0887	16.1638*
1	−1,064.64	87.885	36	0.000	0.331592	15.9229	16.2832	16.8096
2	−1,018.74	91.798	36	0.000	0.288371	15.7805	16.4496	17.4271
3	−915.339	206.81	36	0.000	0.110108*	14.8106*	15.7887*	17.2173
4	−884.081	62.516*	36	0.004	0.119474	14.8789	16.1657	18.04

Table IV.
The test of lag selection by Lütkepohl (2005)

Note: *Means the suggestions of lag selection order from each method

Anderson (1951), Johansen (1998), Gonzalo and Pitarakis (1998) and Aznar and Salvador (2002) for producing statistics which are used to determine the number of co-integrating equations in a vector error-correction model (VECM). Based on our characteristics of the data set with the selected lags, we extract the results in Table V.

The results shown in the table above indicates that there is a co-integrating relationship between any of two pairs (for example, these keywords and percentage in a number of new business establishment or capital flow in registered firms). It interprets that a pair of each keyword and number of new business establishment (or chartered capital) does not persist in the long term. Therefore, any change in a number of new business establishments (or the amount of chartered capital) is intuitively triggered by the short-term shock. Hence, it leads us to use the VAR estimation for further investigation.

4.4 Vector auto regression estimation and granger causality analysis

In previous studies, Nasir *et al.* (2019) and Luu Duc Huynh (2019) also used VAR and VAR Granger to see the causal relationship in financial aspect. However, to determine the managerial direction of causality, we performed the Granger causality test and the results presented in Table VI.

The results of Granger causality test showed that there is no keyword, which has strong evidence to cause the new business establishment during the period from the beginning of 2016 to October 2018. This means that the Granger Causality cannot be used to predict the number of new business registration in Vietnam. However, we also find out something interesting from this approach. The number of new companies established in Vietnam has a significant relationship at 1 per cent to cause the value of searching these keywords "entrepreneurship" (for both English and Vietnamese), "new business", "start-up". In addition, this test has also explained the behavior of participants, who pay attention to entrepreneurship, seems to not look for only one keyword. For example, the searching keyword "entrepreneur" has strong evidence to predict the actions of finding keyword "start-up" and vice versa on Google Search Engine at significance level 5 per cent. To perform the further test, we still employ Granger Causality for chartered capital registered to answer whether these keywords cause a change for capital (Table VII).

Rank	Parms	LL	Eigenvalue	Statistic	Value
0	78	−1,084.03		324.5915	94.15
1	89	−1,007.76	0.66364	172.0509	68.52
2	98	−976.189	0.36301	108.9113	47.21
3	105	−955.218	0.25888	66.9678	29.68
4	110	−935.105	0.24973	26.7431	15.41
5	113	−927.487	0.10313	11.5056	3.76
6	114	−921.734	0.0789		

Notes: Sample: 2016 week 5 – 2018 week 40 (with the lags 3); *, ** and *** significant at 10%, 5% and 1% levels, respectively; because of ENTVN being non-stationary in significance level 1%, we employ further process for this variable by taking the first difference of ENTVN, which is considered as one typical way to deal with stochastic trends (unit root); this test is used for ENT, NB, D.ENTVN, SU, PERNUM, PERCAP variables; the Johansen test is used with Maximum Likelihood estimator of the parameters of a co-integrating VECM with function as follows $\Delta y_t = \alpha\beta' y_{t-1} + \sum_{t=1}^{p-1} \Gamma_i \Delta y_{t-1} + \epsilon_t$. Denoted y is (K x 1) vector of I(1) variables,α and β are (K x r) parameter matrices with rank r < K, $\Gamma_1 \ldots \Gamma_{p-1}$ are (K x K) matrices of parameters and ϵ_t is a (K x 1) vector of normally distributed errors that is serially uncorrelated (Anderson, 1951) and (Johansen 1998)

Table V.
Johansen tests for co-integration for all variables

Null hypothesis	χ^2	Results
PERNUM does not indicate Granger Causality with ENT	2.3605	Fail to reject
PERNUM does not indicate Granger Causality with NB	1.096	Fail to reject
PERNUM does not indicate Granger Causality with D.ENTVN	8.2081**	*Reject*
PERNUM does not indicate Granger Causality with SU	5.9447	Fail to reject
PERNUM does not indicate Granger Causality with all	27.025***	*Reject*
ENT does not indicate Granger Causality with PERNUM	2.7838	Fail to reject
ENT does not indicate Granger Causality with NB	0.93254	Fail to reject
ENT does not indicate Granger Causality with D.ENTVN	2.7848	Fail to reject
ENT does not indicate Granger Causality with SU	9.8871**	*Reject*
ENT does not indicate Granger Causality with all	18.214	Fail to reject
NB does not indicate Granger Causality with PERNUM	3.0498	Fail to reject
NB does not indicate Granger Causality with ENT	0.46674	Fail to reject
NB does not indicate Granger Causality with D.ENTVN	8.271**	*Reject*
NB does not indicate Granger Causality with SU	4.9168	Fail to reject
NB does not indicate Granger Causality with all	16.433	Fail to reject
D.ENTVN does not indicate Granger Causality with PERNUM	2.3841	Fail to reject
D.ENTVN does not indicate Granger Causality with ENT	3.6176	Fail to reject
D.ENTVN does not indicate Granger Causality with NB	5.6319	Fail to reject
D.ENTVN does not indicate Granger Causality with SU	5.2743	Fail to reject
D.ENTVN does not indicate Granger Causality with all	16.589	Fail to reject
SU does not indicate Granger Causality with PERNUM	4.1606	Fail to reject
SU does not indicate Granger Causality with ENT	10.116**	*Reject*
SU does not indicate Granger Causality with NB	6.0134	Fail to reject
SU does not indicate Granger Causality with D.ENTVN	8.8352	Fail to reject
SU does not indicate Granger Causality with all	27.515***	*Reject*

Note: *, ** and *** significant at 10%, 5% and 1% levels, respectively

Table VI.
Granger causality for percentage in a change of new firm establishment

Interestingly, there is also no keyword, which causes a change in the amount of chartered capital registered in the research period. However, it is once again to validate the behavioral of users searching the keyword "entrepreneur" and "start-up" is strongly correlated by two sides from Appendix. To clarify the keywords' impact on a change of new business establishment or an amount of money invested; we use a further set of quantitative techniques to investigate the insights here.

4.5. Threshold regression estimation

Threshold regression allows the model with their coefficients to differ across regions, which is determined by an above or below the threshold. This approach is proposed by Tong (1983, 1990), Hansen (2011). Our estimation is mainly used for conditional least squares to estimate the parameters of the threshold regression model.

We attract our attention to the effect of a change in the new business establishment under the value of keywords searched. In our model, we assume that the Vietnamese Government sets the threshold for the absolute number of firms born. We use the absolute figures for new companies established in Vietnam as one threshold (or two regions). Thus, the model may be written as:

$$\begin{aligned} \mathrm{PERNUM}_t = {} & \delta_{10} + \delta_{11}\,\mathrm{PERNUM}_{t-1} + \delta_{12}\,\mathrm{ENT}_t + \delta_{13}\,\mathrm{NB}_t + \delta_{14}\,\mathrm{D.ENTVN}_t \\ & + \delta_{14}\,\mathrm{SU}_t + \epsilon_t \ \text{if} - \infty < NUM < \gamma \end{aligned} \tag{6}$$

Null hypothesis	χ^2	Results
PERCAP does not indicate Granger Causality with ENT	2.6267	Fail to reject
PERCAP does not indicate Granger Causality with NB	3.328	Fail to reject
PERCAP does not indicate Granger Causality with D.ENTVN	3.6538	Fail to reject
PERCAP does not indicate Granger Causality with SU	2.3063	Fail to reject
PERCAP does not indicate Granger Causality with all	10.77	Fail to reject
ENT does not indicate Granger Causality with PERCAP	1.1722	Fail to reject
ENT does not indicate Granger Causality with NB	1.1308	Fail to reject
ENT does not indicate Granger Causality with D.ENTVN	1.7601	Fail to reject
ENT does not indicate Granger Causality with SU	9.98**	*Reject*
ENT does not indicate Granger Causality with all	16.428	Fail to reject
NB does not indicate Granger Causality with PERCAP	3.5342	Fail to reject
NB does not indicate Granger Causality with ENT	0.13862	Fail to reject
NB does not indicate Granger Causality with D.ENTVN	5.525	Fail to reject
NB does not indicate Granger Causality with SU	5.9225	Fail to reject
NB does not indicate Granger Causality with all	16.963	Fail to reject
D.ENTVN does not indicate Granger Causality with PERCAP	5.6833	Fail to reject
D.ENTVN does not indicate Granger Causality with ENT	4.0682	Fail to reject
D.ENTVN does not indicate Granger Causality with NB	6.1207	Fail to reject
D.ENTVN does not indicate Granger Causality with SU	4.4209	Fail to reject
D.ENTVN does not indicate Granger Causality with all	20.217*	*Reject*
SU does not indicate Granger Causality with PERCAP	3.3163	Fail to reject
SU does not indicate Granger Causality with ENT	11.782***	*Reject*
SU does not indicate Granger Causality with NB	7.5531	Fail to reject
SU does not indicate Granger Causality with D.ENTVN	6.705*	Fail to reject
SU does not indicate Granger Causality with all	26.533***	*Reject*

Note: *, ** and *** *significant at 10%, 5% and 1% levels, respectively*

Table VII. Granger causality for percentage in a change of capital registered

$$\mathrm{PERNUM}_t = \delta_{20} + \delta_{21}\,\mathrm{PERNUM}_{t-1} + \delta_{22}\,\mathrm{ENT}_t + \delta_{23}\,\mathrm{NB}_t + \delta_{24}\,\mathrm{D.ENTVN}_t + \delta_{24}\,\mathrm{SU}_t + \epsilon_t \ \text{ if } \ \gamma < NUM < +\infty \tag{7}$$

$$\mathrm{PERCAP}_t = \delta_{30} + \delta_{31}\,\mathrm{PERCAP}_{t-1} + \delta_{32}\,\mathrm{ENT}_t + \delta_{33}\,\mathrm{NB}_t + \delta_{34}\,\mathrm{D.ENTVN}_t + \delta_{34}\mathrm{SU}_t + \epsilon_t \ \text{ if } -\infty < CAP < \gamma \tag{8}$$

$$\mathrm{PERCAP}_t = \delta_{40} + \delta_{41}\,\mathrm{PERCAP}_{t-1} + \delta_{42}\,\mathrm{ENT}_t + \delta_{43}\,\mathrm{NB}_t + \delta_{44}\,\mathrm{D.ENTVN}_t + \delta_{44}\,\mathrm{SU}_t + \epsilon_t \ \text{ if } \ \gamma < NUM < +\infty \tag{9}$$

The results for threshold regression are illustrated in Table VIII.

This table reports the coefficients in threshold regression and the corresponding Sum of Squared Residuals (SSR). We only choose a single threshold, which contributes to the most value in minimizing the SSR. The estimated threshold of 1,602 firms and 35,010m VND (Vietnam dong, Vietnamese currency) split the sample into two regions. The region 1 corresponds to the portion of the sample in which a quantity of new companies is less than or equal to 1,602 firms and an amount of money invested to register new ones is under 35,010m VND. Region 2 corresponds to the portion of the sample in which two aforementioned value is greater than 1,602 and 35,010, respectively.

Y_1 = PERNUM	Coef	Y_2 = PERCAP	Coef
	Region 1		Region 1
L1.PERNUM	−4.055946*** [0.6895654]	L1.PERCAP	−0.1062225* [0.0612877]
ENT	1.019179*** [0.242405]	ENT	0.0763683 [0.0824977]
NB	2.179222*** [0.2884519]	NB	0.029928 [0.0871885]
D1.ENTVN	−1.177853*** [0.2671696]	D1.ENTVN	−0.1658348 [0.1543781]
SU	−2.354959*** [0.2797288]	SU	0.0066573 [0.0903958]
Cons	−0.3830465 [0.249589]	Cons	0.088861 [0.0841756]
	Region 2		Region 2
L1.PERNUM	0.0416155 [0.0566236]	L1.PERCAP	2.347929 [0.8533764]
ENT	0.0137247 [0.0567305]	ENT	2.631785*** [0.4544781]
NB	0.0189777 [0.0574766]	NB	0.1080512 [0.2463264]
D1.ENTVN	0.0199506 [0.1165031]	d1.ENTVN	0.4319265 [0.6552301]
SU	−0.0207872 [0.0620641]	SU	−2.532224*** [0.3597332]
Cons	0.0602966 [0.0569238]	Cons	2.675673*** [0.317829]
Order 1 – Threshold	1,602	Order 1 – Threshold	35,010
Sum of Squared Residuals	52.9929	Sum of Squared Residuals	114.8411

Note: *, ** and *** significant at 10%, 5% and 1% levels, respectively

Table VIII.
Results of threshold regression with two regions with γ_1 (number of firms), γ_2 (amount of chartered capital)

When it comes to the coefficients of the number of new companies established, the lower region witnesses the strong relationship by independent variables by the lag of a percentage change in new firm figures, keywords such as "entrepreneur", "new business", the first difference in Vietnamese terminology for "entrepreneur" and "start-up". There are three factors, which cause a negative impact on the number of new companies being initially run in Vietnam, such as L.PERNUM, D.ENTVN and SU. However, the variables ENT and NB positively trigger an increase in new businesses in Vietnam. The lag of L1.pernum will contribute to an increase in the percentage of new business at the significance level of 1 per cent. When users observe the booming of new enterprises in this period, they might take into consideration to register the new ones in the following period. Therefore, if there is an increase in the former week, the next week experiences a decrease from users, which are supposed to carefully observe the markets.

Regards the keywords, we figure out that there are some interesting points. When the users find more and more frequency of "entrepreneur" in English, the number of new businesses will increase at the significance level 1 per cent. In contrast, the possibilities of searching "entrepreneur' in Vietnamese cause a decrease in a number of new companies at the significance level 1 per cent. It can be explained by participation of both foreigners and locals, who are staying in Vietnam for searching these keywords by Google statistical figures. When the foreigners search for information regarding entrepreneurship in English, they might find potential opportunities. Thus, they come into the decision to register a new business in a dynamic market in Vietnam. Hence, we can interpret that the entrepreneurial information in English can be predicted for a rising number of new firms in Vietnam, which is considered as a risk-taking for those finding by English keyword of the entrepreneur. Meanwhile, the Vietnamese information related to entrepreneur negatively influences on statistical figures of new companies established. It is noted that the findings are under the lower region and Searching-Google-based-on-Vietnam-geography.

There are some differences between results from "start-up" and "new-business" keyword in explaining the number of firms in Vietnam. "New-business" is positive co-efficient

whereas "start-up" cause a negative impact on new firms established. Our explanation for this case is to rely on the stage of establishing a new business. When users find information with keyword "start-up", this means the beginning stage of market research or a collection of relevant concerns. However, if users use keyword "new business", this is over the final stage of registration or the firms' operations. Therefore, the information for "new business" encourages the users to run the new firms, which supports a rising percentage in new enterprises. However, users find difficulties when initially searching for "start-up'; therefore, they might be risk-averse behaved regarding new firms establishment. These findings are workable for the lower region.

Surprisingly, the higher region experience no keywords, which significantly influence a number of new companies, registered in Vietnam at any significance level. Finally, our threshold estimation for interpretation is 1,602 companies. According to Appendix 2, our estimated threshold witnessed an increase after declining between 1,000 and 2,000 companies in the researched period. Then, these searched keywords might suggest some implications for those who wish to commence their business in terms of information searching. Turning to the amount of capital used for registering new companies, there are also two regions with a threshold defined as 35,010m VND. In the context of the lower region, only the lag of variable "PERCAP" has a negative impact on an amount of chartered capital invested in new business at significance level 1 per cent. When it comes to keywords, the keyword "entrepreneur" in English demonstrates a rise for "PERCAP" whereas "START-UP" causes a decline at significance level 1 per cent. The explanation for this phenomenon is about the users who get used to with English and show strong motivation to do business in Vietnam when searching in English. However, the keyword "start-up" is quite popular for those who just want to initially penetrate in the market.

5. Implications and conclusion

There are various studies, which investigate the relationship between the Google keywords and the other elements such as stock returns, unemployment rates. This paper contributes a piece of new empirical evidence regarding Google keywords and the number of firms and an amount of registered capital in the context of Vietnam by threshold regression.

Therefore, the study presents some meaningful findings. *Firstly*, the Vietnamese government and executives need to be careful to search the information raising from keywords "entrepreneur" (in Vietnamese) and '*start-up*', which leads to some information resulting in a decline the number of firms and an amount of capital. Instead, the keywords "*new business*" and "entrepreneur' have a strong motivation to create more firms and to foster money for the whole economy. *Secondly*, there is a threshold as 1,602 companies and 35,010 VND, which presents the differences in keywords" influence. In the lower regime of the number of companies, most keywords contribute to a quantity change of firms.

However, it does not persist in the higher region. In contrast, the higher region for an amount of money is influenced by keywords. *Finally*, this study suggests that the process of company establishment is not caused by any long-term shocks. In the short run, there is a small change among these keywords by shifting from one to another one. Significantly, the threshold is an important element, which results in any changes in the number of new businesses or an amount of capital.

References

Acs, Z.J., Desai, S. and Hessels, J. (2008), "Entrepreneurship, economic development and institutions", *Small Business Economics*, Vol. 31 No. 3, pp. 219-234.

Anderson, T.W. (1951), "Estimating linear restrictions on regression coefficients for multivariate normal distributions", *The Annals of Mathematical Statistics*, Vol. 22 No. 3, pp. 327-351.

Askitas, N. and Zimmermann, K.F. (2009), "Google econometrics and unemployment forecasting", *Applied Economics Quarterly*, Vol. 55 No. 2, pp. 107-120.

Aznar, A. and Salvador, M. (2002), "Selecting the rank of the cointegration space and the form of the intercept using an information criterion", *Econometric Theory*, Vol. 18 No. 4, pp. 926-947.

Baker, S., Bloom, N. and Davis, S.J. (2016), "Measuring economic policy uncertainty", *The Quarterly Journal of Economics*, Vol. 131 No. 4, pp. 1593-1636.

Bijl, L., Kringhaug, G., Molnar, P. and Sandvik, E. (2016), "Google searches and stock returns", *International Review of Financial Analysis*, Vol. 45 No. 13, pp. 150-156.

Bollen, J., Mao, H. and Zeng, X. (2011), "Twitter mood predicts the stock market", *Journal of Computational Science*, Vol. 2 No. 1, pp. 1-8.

Carnes, C.M., Gilstrap, F.E., Hitt, M.A., Ireland, R.D., Matz, J.W. and Woodman, R.W. (2019), "Transforming a traditional research organization through public entrepreneurship", *Business Horizons*, Vol. 62 No. 4, pp. 437-449.

Courtois, C., Slechten, L. and Coenen, L. (2018), "Challenging Google search filter bubbles in social and political information: disconforming evidence from a digital methods case study", *Telematics and Informatics*, Vol. 35 No. 7, pp. 2006-2015.

Da, Z., Engelberg, J. and Gao, P. (2011), "In search of attention", *The Journal of Finance*, Vol. 66 No. 5, pp. 1461-1499.

Davis, P.S. and Harveston, P.D. (2000), "Internationalization and organizational growth: the impact of internet usage and technology involvement among entrepreneur-led family businesses", *Family Business Review*, Vol. 13 No. 2, pp. 107-120.

Dickey, D.A. and Fuller, W.A. (1979), "Distribution of the estimators for autoregressive time series with a unit root", *Journal of the American Statistical Association*, Vol. 74 No. 366, pp. 427-431.

Gifford, S. (1992), "Allocation of entrepreneurial attention", *Journal of Economic Behavior and Organization*, Vol. 19 No. 3, pp. 265-284.

Gonzalo, J. and Pitarakis, J.Y. (1998), "Specification via model selection in vector error correction models", *Economics Letters*, Vol. 60 No. 3, pp. 321-328.

Granger, C.W.J. (1969), "Investigating causal relations by econometric models and cross-spectral methods", *Econometrica*, Vol. 37 No. 3, pp. 424-438.

Gwilym, O.A., Hasan, I., Wang, Q. and Xie, R. (2016), "In search of concepts: the effects of speculative demand on stock returns", *European Financial Management*, Vol. 22 No. 3, pp. 427-449.

Hansen, B.E. (2000), "Sample splitting and threshold estimation", *Econometrica*, Vol. 68 No. 3, pp. 575-603.

Hansen, B.E. (2011), "Threshold autoregression in economics", *Statistics and Its Interface*, Vol. 4 No. 2, pp. 123-127.

Huber, G.P. (1991), "Organizational learning: the contributing processes and the literatures", *Organization Science*, Vol. 2 No. 1, pp. 88-115.

Johansen, S. (1988), "Statistical analysis of cointegration vectors", *Journal of Economic Dynamics and Control*, Vol. 12 Nos 2/3, pp. 231-254.

Kirihata, T. (2018), "Japanese government venture capital: what should we know?", *Asia Pacific Journal of Innovation and Entrepreneurship*, Vol. 12 No. 1, pp. 14-31.

Kornish, L.J. and Ulrich, K.T. (2011), "Opportunity spaces in innovation: empirical analysis of large samples of ideas", *Management Science*, Vol. 57 No. 1, pp. 107-128.

Lee, Y. and Chun, H. (2017), "Nation image and its dynamic changes in Wikipedia", *Asia Pacific Journal of Innovation and Entrepreneurship*, Vol. 11 No. 1, pp. 38-49.

Li, Q., Maggitti, P.G., Smith, K.G., Tesluk, P.E. and Katila, R. (2013), "Top management attention to innovation: the role of search selection and intensity in new product introductions", *Academy of Management Journal*, Vol. 56 No. 3, pp. 893-916.

Lopez, L. and Weber, S. (2017), "Testing for granger causality in panel data", *The Stata Journal: Promoting Communications on Statistics and Stata*, Vol. 17 No. 4, pp. 972-984.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer, New York, NY.

Luu Duc Huynh, T. (2019), "Spillover risks on cryptocurrency markets: a look from VAR-SVAR granger causality and student'st copulas", *Journal of Risk and Financial Management*, Vol. 12 No. 2, pp. 52.

Nasir, M.A., Huynh, T.L.D., Nguyen, S.P. and Duong, D. (2019), "Forecasting cryptocurrency returns and volume using search engines", *Financial Innovation*, Vol. 5 No. 1, pp. 2-10.

Neri, K., Katarına Lucivjanskă, P., Molnar, R. and Villa (2018), "Google' searches and stock market activity: evidence from Norway", *Finance Research Letters*, Vol. 28 No. 3, pp. 208-220.

Newey, W.K. and West, K.D. (1987), "A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix", *Econometrica*, Vol. 55 No. 3, pp. 703-708.

Nga, J.K.H. and Shamuganathan, G. (2010), "The influence of personality traits and demographic factors on social entrepreneurship start up intentions", *Journal of Business Ethics*, Vol. 95 No. 2, pp. 259-282.

Phillips, P.C.B. and Perron, P. (1988), "Testing for a unit root in time series regression", *Biometrika*, Vol. 75 No. 2, pp. 335-346.

Ribeiro, A.T.V.B. and Plonski, G.A. (2019), "Grassroots movements: new gears in the engine of Brazilian university entrepreneurship ecosystem", *In Startups and Innovation Ecosystems in Emerging Markets*, Palgrave Macmillan, Cham, pp. 179-201.

Scott, A.J. (2006), "Entrepreneurship, innovation and industrial development: geography and the creative field revisited", *Small Business Economics*, Vol. 26 No. 1, pp. 1-24.

Statista (2019), "Number of internet users in vietnam from 2017 to 2023 (in millions)", available at: www.statista.com/statistics/369732/internet-users-vietnam/ (accessed 16 June 2019).

Thaler, R.H. (1987), "Seasonal movements in security prices I: the January effect", *Journal of Economic Perspectives*, Vol. 1 No. 1, pp. 197-201.

Tong, H. (1983), *Threshold Models in Non-Linear Time Series Analysis*, Springer, New York, NY.

Tong, H. (1990), "Non-linear time series: a dynamical system approach", Oxford University Press, New York, NY.

Xiang, Z. and Gretzel, U. (2010), "Role of social media in online travel information search", *Tourism Management*, Vol. 31 No. 2, pp. 179-188.

Yan, Y. and Guan, J. (2019), "Entrepreneurial ecosystem, entrepreneurial rate and innovation: the moderating role of internet attention", *International Entrepreneurship and Management Journal*, Vol. 15 No. 2, pp. 625-650.

Yang, X., Sun, S.L. and Zhao, X. (2019), "Search and execution: examining the entrepreneurial cognitions behind the lean startup model", *Small Business Economics*, Vol. 52 No. 3, pp. 667-679.

Further reading

Al-Mamun, A. and Fazal, S.A. (2018), "Effect of entrepreneurial orientation on competency and micro-enterprise performance", *Asia Pacific Journal of Innovation and Entrepreneurship*, Vol. 12 No. 3, pp. 379-398.

D'Amuri, F. and Marcucci, J. (2017), "The predictive power of google searches in forecasting US unemployment", *International Journal of Forecasting*, Vol. 33 No. 4, pp. 801-816.

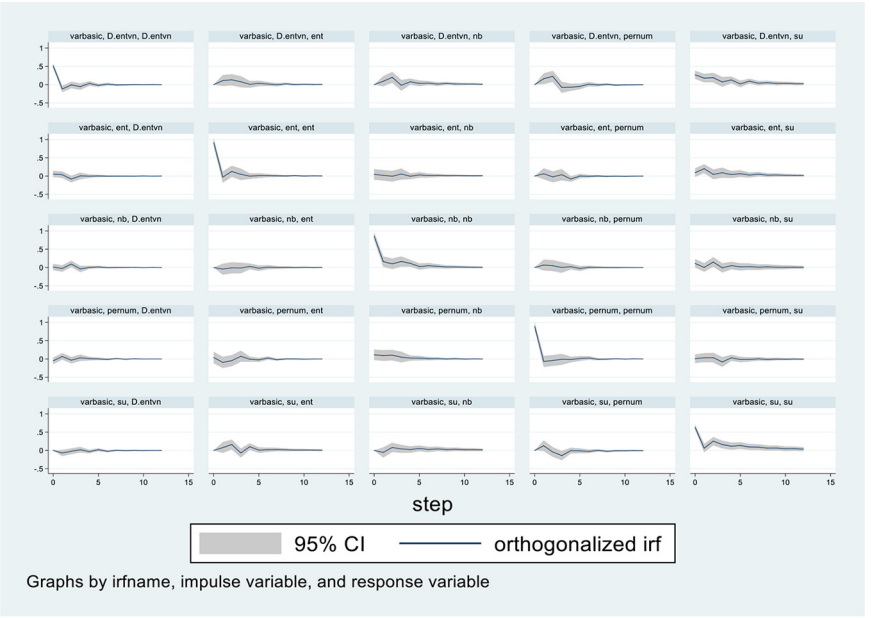
Shane, S. (2009), "Why encouraging more people to become entrepreneurs is bad public policy", *Small Business Economics*, Vol. 33 No. 2, pp. 141-149.

Gonzalo, J. and Pitarakis, J.Y. (2002), "Estimation and model selection based inference in single and multiple threshold models", *Journal of Econometrics*, Vol. 110 No. 2, pp. 319-352.

Ivanov, V. and Kilian, L. (2005), "A practitioner's guide to Lag-Order selection for vector autoregressions", *Studies in Nonlinear Dynamics and Econometrics*, Vol. 9 No. 1, pp. 1-34.

Tran, A.T. and Von Korflesch, H. (2016), "A conceptual model of social entrepreneurial intention based on the social cognitive career theory", *Asia Pacific Journal of Innovation and Entrepreneurship*, Vol. 10 No. 1, pp. 17-38.

Appendix



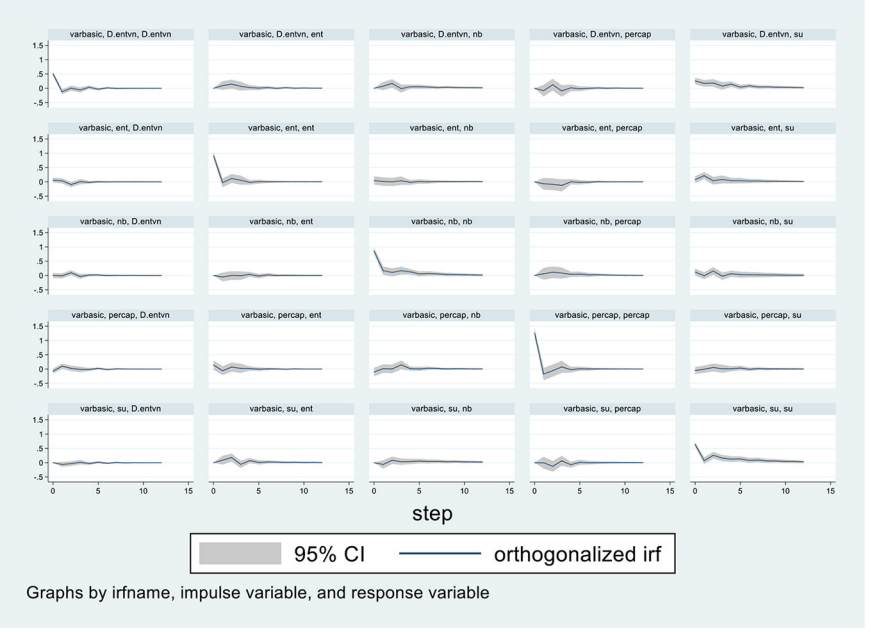


Figure A1. The impulse-response function analysis

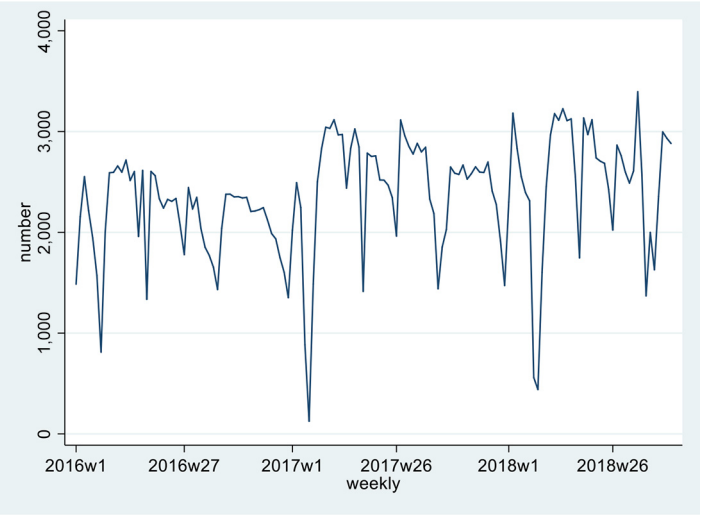


Figure A2.
The number of new firms in Vietnam illustrated by weekly

About the author

Toan Luu Duc Huynh is working as a Research Assistant at Chair of Behavioral Finance, WHU – Otto Beisheim School of Management, Vallendar, Germany. Toan Luu Duc Huynh can be contacted at: toanhld@ueh.edu.vn and toan.huynh@whu.edu