Industrial Management & Data Systems

Integer-valued GARCH processes for Apple technology analysis

Jong-Min Kim, Sunghae Jun,

Integer-valued GARCH processes for Apple technology analysis

Jong-Min Kim
Department of Statistics, University of Minnesota at Morris, Morris, Minnesota, USA, and
Sunghae Jun
Department of Statistics, Cheongju University, Cheongju, Republic of Korea and Division of Industrial Management Engineering, Korea University, Seoul, Republic of Korea

Received 20 January 2017
Revised 19 March 2017
Accepted 31 March 2017

Abstract

Purpose – The keywords from patent documents contain a lot of information of technology. If we analyze the time series of keywords, we will be able to understand even more about technological evolution. The previous researches of time series processes in patent analysis were based on time series regression or the Box-Jenkins methodology. The methods dealt with continuous time series data. But the keyword time series data in patent analysis are not continuous, they are frequency integer values. So we need a new methodology for integer-valued time series model. The purpose of this paper is to propose modeling of integer-valued time series for patent analysis.

Design/methodology/approach – For modeling frequency data of keywords, the authors used integer-valued generalized autoregressive conditional heteroskedasticity model with Poisson and negative binomial distributions. Using the proposed models, the authors forecast the future trends of target keywords of Apple in order to know the future technology of Apple.

Findings – The authors carry out a case study to illustrate how the methodology can be applied to real problem. In this paper, the authors collect the patent documents issued by Apple, and analyze them to find the technological trend of Apple company. From the results of Apple case study, the authors can find which technological keywords are more important or critical in the entire structure of Apple's technologies.

Practical implications – This paper contributes to the research and development planning for producing new products. The authors can develop and launch the innovative products to improve the technological competition of a company through complete understanding of the technological keyword trends.

Originality/value – The retrieved patent documents from the patent databases are not suitable for statistical analysis. So, the authors have to transform the documents into structured data suitable for statistics. In general, the structured data are a matrix consisting of patent (row) and keyword (column), and its element is an occurred frequency of a keyword in each patent. The data type is not continuous but discrete. However, in most researches, they were analyzed by statistical methods for continuous data. In this paper, the authors build a statistical model based on discrete data.

Keywords Patent analysis, Apple keywords, Integer-values time series model,
Poisson and negative binomial distributions

Paper type Research paper

1. Introduction

Many case studies on Apple's technological innovation have been fulfilled in many academic and industrial fields (Funk, 2011; Arruda-Filho *et al.*, 2010; Arruda-Filho and Lennon, 2011; West and Mace, 2010; Halal, 2013). The research studies focused on Apple's technological evolution. Apple is one of the global innovative companies leading the smartphone market (Nam *et al.*, 2015; Hung *et al.*, 2013; Wonglimpiyarat, 2005). A number of business schools have studied on the technological development and innovation of Apple for other companies' technological innovations. Many companies have pursuit the technological innovative strategies of Apple. To know the technological innovation of Apple, we have to analyze Apple's technologies. Apple's patents contain more information of developed technologies of Apple than other resources such as papers or articles because the patent system protects the exclusive right of researched and developed technologies for

Industrial Management & Data Systems
Vol. 117 No. 10, 2017
pp. 2381-2399

0263-5577
DOI 10.1108/IMDS-01-2017-0023

the inventors (Hunt *et al.*, 2007). In this paper, we use entire patents applied by Apple for Apple's technology analysis. Jun and Park (2013) analyzed Apple's patents using statistical methods and the social network analysis for examining Apple's technological innovation (Jun and Park, 2013). In addition, Kim and Jun (2015) proposed graphical causal inference and the copula regression model for the keyword analysis of Apple's patents (Kim and Jun, 2015). They used advanced statistical inference and visualization to construct Apple's technology map. In this paper, we use technological keywords of patent data. The keywords from Apple's patent documents also contain diverse and complete information of Apple's technology. In addition, if we analyze the time series data of the keywords, we will be able to understand even more about the technological evolution of Apple. Guidolin and Guseo (2014) performed the seasonality modeling for Apple's innovative diffusion (Guidolin and Guseo, 2014).

It is meaningful to analyze patent data with the passage of time (Guidolin and Guseo, 2014; Hong *et al.*, 2016; Lakka *et al.*, 2013). In many studies on technology analysis of company as well as Apple, diverse time series analysis models were used to understand the technological trends of a company (Jun and Uhm, 2010; Park and Jun, 2012; Jun, 2013; Park *et al.*, 2016). These research studies of time series processes in patent analysis were based on time series regression or the Box-Jenkins methodology. The methods dealt with continuous time series data. But the keyword time series data in patent analysis are not continuous, they are integer values. So we need another methodology for the integer-valued time series analysis for the patent keyword analysis. This paper discusses the modeling of integer-valued time series with Apple's keywords. For modeling count data of the keywords, we use the integer-valued generalized autoregressive conditional heteroskedasticity (INGARCH) model with Poisson distribution and negative binomial distribution (Bollerslev, 1986; Christou and Fokianos, 2014; Davis and Wu, 2009; Engle, 1982; Ferland *et al.*, 2006; Fokianos and Fried, 2010, 2012). The results from numerical studies indicate that the negative binomial INGARCH (INGARCH-NB) model performs better than the Poisson INGARCH (INGARCH-Pois) model; and the INGARCH-NB model with covariate keyword performs better than the INGARCH-NB model without covariate keyword. In particular, using the Apple's keywords by text mining, we show that the negative binomial integer-valued autoregressive conditional heteroscedasticity (INARCH-NB) model has smaller Akaike information criterion (AIC) value than any other models. So using the INARCH(1)-NB model, we forecast our target Apple Keyword with other high correlated Apple Keyword in order to know the future technology trend.

We organized our paper as follows. Section 2 introduces the way to extract keywords from patent data. We propose count time series models for Apple's keywords in Section 3. In Section 4, we illustrate the examples with Apple's keywords by text mining. Finally, we show our conclusions in Section 5.

2. Keyword extraction form Apple's patent documents

In this paper, we use technological keywords from retrieved Apple's patent documents. First of all, we retrieve entire patents applied by Apple in the world. We use the patent databases of WIPS Corporation and the United States Patent and Trademark Office (WIPSON, 2014; USPTO, 2015). Figure 1 shows the keyword extraction, structured data matrix, and integer-valued time series modeling.

In this paper, we use R data language and its packages for data preprocessing and analysis (Feinerer and Hornik, 2014; R Core Team, 2014). We build two structured data matrices which are patent-keyword matrix (PKM) and year-keyword matrix (YKM). Using the PKM, we make the correlation structure between Apple's technological keywords, and we perform time series analysis using the YKM. So, we construct integer-valued time series modeling using the PKM and YKM.

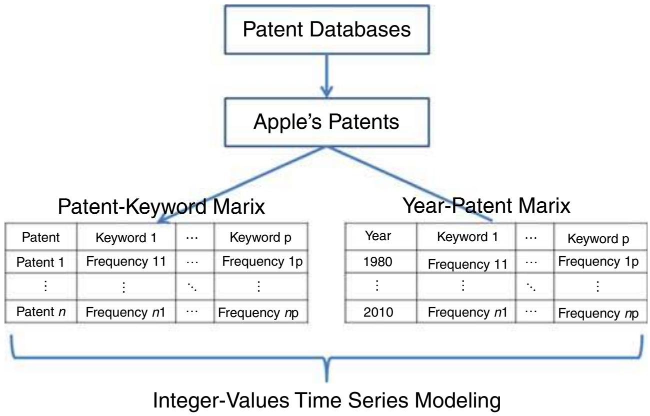

Figure 1. Text mining procedure of Apple's keywords

3. Count time series models for Apple keywords

Engle (1982) and Bollerslev (1986) proposed the autoregressive conditional heteroskedasticity (ARCH) and the generalized autoregressive conditional heteroskedasticity (GARCH) models to explain conditionally heteroscedastic models (Engle, 1982; Bollerslev, 1986). There is a long list of variations of the GARCH models that also consider the asymmetricity. For more details about the asymmetric GARCH models, refer Engle and Ng (1993). The Generalized ARCH (GARCH) (p, q) model of Bollerslev (1986) is expressed as:

$$X_t|F_{t-1} \sim N(0, h_t), h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j},$$

where $\alpha_0 > 0$, $\alpha_i \geqslant 0$, $\beta_j \geqslant 0$, $i = 1, \ldots, p$, $j = 1, \ldots, q$, $p \geqslant 1$, $q \geqslant 1$. In addition, F_{t-1} is the history of the process up to time $t-1$. Since Zeger (1998) proposed the regression model for time series of counts with an analysis of trends in the US polio incidence since 1970, the model for analyzing time series of counts has been getting more attention among researchers. In recent years, many diverse models for analyzing time series of counts have been proposed and have been applied to the wide range of practical applications such as a biological experiment, epidemiology and finance. The detailed examples with models for time series of counts can be found in Jung *et al.* (2006) and Davis and Wu (2009). We introduce the reader the integer-valued GARCH model which is the most flexible time series model of count data in this section and then apply the model to Apple's keywords by text mining in the following section. The paper is the first paper to apply the time series model for count data to Apple's keywords by text mining. It will be the most significant contribution to text mining research areas.

3.1 Integer-valued GARCH models

Ferland *et al.* (2006) proposed the integer-valued GARCH (p, q) with Poisson distribution which will be denoted as the INGARCH (p, q) process in this paper. Let $\{X_t\}$ be a time series of counts. We assume that, conditional on the past information, the random variables $X_1, X_2, \ldots, X_n$ are independent, and the conditional distribution of X_t is specified by a Poisson distribution, i.e., $X_t|F_{t-1}$~Poisson (λ_t), satisfying:

$$\Pr(X_t = x_t|F_{t-1}) = \frac{\exp(-\lambda_t)\lambda_t^{y_t}}{y_t!}, \quad y_t = 0, 1, \ldots. \tag{1}$$

with:

$$\lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}, \tag{2}$$

where we define F_{t-1} to be the history of the process up to time $t-1$, $\alpha_0 > 0$, $\alpha_i \geqslant 0$, $\beta_j \geqslant 0$, $i = 1, \ldots, p$, $j = 1, \ldots, q$, $p \geqslant 1$, $q \geqslant 1$. When $q = 0$, the INGARCH (p, q) which is the integer-valued GARCH (p, q) with Poisson distribution becomes the integer-valued ARCH (p) with Poisson distribution denoted by INARCH (p). Then (2) becomes:

$$\lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}. \tag{3}$$

The Poisson distribution provides a standard framework for the analysis of count data, but the requirement that the variance should equal the mean is often too restrictive in practice. Frequently data are overdispersed, with the variance greater than the mean. So Zhu (2011) proposed the negative binomial integer-valued GARCH (p, q) which will be denoted as the INGARCH (p, q) – NB process in this paper. Let $\{X_t\}$ be a time series of counts. We assume that, conditional on the past information, the random variables $X_1, X_2, \ldots, X_n$ are independent, and the conditional distribution of X_t is specified by a negative binomial distribution, i.e., $X_t|F_{t-1} \sim \mathrm{NB}(r, p_t)$, satisfying:

$$\Pr(X_t = x_t | F_{t-1}) = \binom{x_t + r - 1}{r - 1} \left(\frac{1}{1+p_t}\right)^r \left(\frac{p_t}{1+p_t}\right)^{y_t}, \tag{4}$$

with:

$$\lambda_t = \frac{1-p_t}{p_t} = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}, \tag{5}$$

where we define F_{t-1} to be the history of the process up to time $t-1$, $\alpha_0 > 0$, $\alpha_i \geqslant 0$, $\beta_j \geqslant 0$, $i = 1, \ldots, p, j = 1, \ldots, q, p \geqslant 1, q \geqslant 1$. When $q = 0$, the INGARCH (p, q)-NB which is the integer-valued GARCH (p, q) with negative binomial distribution becomes the integer-valued ARCH (p) with Poisson distribution denoted by INARCH (p)-NB. Then (5) becomes:

$$\lambda_t = \frac{1-p_t}{p_t} = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}. \tag{6}$$

3.2 Integer-valued GARCH models with covariates

Fokianos and Fried (2010) proposed the integer-valued GARCH with covariates having an internal effect. Let $\{Y_t\}$ be a time series of counts. We assume that, conditional on the past information, the random variables $Y_1, Y_2, \ldots, Y_n$ are independent, $\{X_t\}$ is some covariate time series and if the conditional distribution of Y_t is specified by a Poisson distribution, i.e.:

$$Y_t | G_{t-1} \sim \mathrm{Poisson}(\omega_t),$$

with:

$$\omega_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{q} \beta_j \omega_{t-j} + \sum_{k=1}^{s} \gamma_j X_{t,r}, \tag{7}$$

where we define G_{t-1} to be the history of the process up to time $t-1$, $\alpha_0 > 0$, $\alpha_i \geqslant 0$, $\beta_j \geqslant 0$, $i = 1, \ldots, p$, $j = 1, \ldots, q$, $p \geqslant 1$, $q \geqslant 1$. When $q = 0$ the INGARCH (p, q) with Poisson distribution given covariates becomes INARCH (p) with covariates. Then (7) becomes:

$$\omega_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{k=1}^{s} \gamma_j X_{t,r}. \tag{8}$$

INGARCH-NB process by Zhu (2011) can be accommodated with time-dependent covariates. Let $\{Y_t\}$ be a time series of counts. We assume that, conditional on the past information, the random variables $Y_1, Y_2, \ldots, Y_n$ are independent, $\{X_t\}$ is some covariate time series and if the conditional distribution of Y_t is specified by a negative binomial distribution, i.e.:

$$Y_t | G_{t-1} \sim \mathrm{NB}(r, \vartheta_t),$$

with:

$$\omega_t = \frac{1-\vartheta_t}{\vartheta_t} = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{q} \beta_j \omega_{t-j} + \sum_{k=1}^{s} \gamma_j X_{t,r}, \tag{9}$$

where we define G_{t-1} to be the history of the process up to time $t-1$, $\alpha_0 > 0$, $\alpha_i \geqslant 0$, $\beta_j \geqslant 0$, $i = 1, \ldots, p$, $j = 1, \ldots, q$, $p \geqslant 1$, $q \geqslant 1$. When $q = 0$, the INGARCH (p, q)-NB with covariates becomes the INARCH (p)-NB with covariates. Then (9) becomes:

$$\omega_t = \frac{1-\vartheta_t}{\vartheta_t} = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{k=1}^{s} \gamma_j X_{t,r}. \tag{10}$$

In addition, the INGARCH-NB model has been developed by Christou and Fokianos (2014) and Liboschik *et al.* (2015).

3.3 Integer-valued GARCH models with covariates

In this paper, we use two measures to evaluate the fitted model. First, we consider the AIC as follows (Joseph, 2011; Akritas, 2016):

$$\mathrm{AIC} = n \log\left(\frac{\mathrm{SSE}}{n}\right) + 2P \tag{11}$$

where the SSE is the error sum of squares, and n and P represent sample size and model parameter. We select the mode with the smallest AIC value for the best model. The Bayes information criterion (BIC) is our second measure for model evaluation which is as follows (Joseph, 2011; Akritas, 2016):

$$\mathrm{BIC} = n \log\left(\frac{\mathrm{SSE}}{n}\right) + P \log(n) \tag{12}$$

Like the AIC measure, the model with the smallest BIC value is the best model.

4. Illustrated examples with Apple's keywords by text mining

Based on the result of Kim and Jun (2015), we have chosen four possible pairs of target variable (Y) and covariate (X) as shown in Table I.

Of course, there exist more relations between the keywords, we selected the most significant ones from all relations between Apple's keywords. In this paper, we select the count time series model for Apple's keywords of "User," "Memory," and "System." In addition, we consider covariate for each keyword; "Interface" is the covariate for "User," "Data" is the covariate for "Memory," "Information" and "Present" are the covariates for "System." Using the keyword structure in Table I, we apply eight different count time series models (four models with covariates and four models without covariates) in this paper to find the best model for forecasting Apple's technological keywords. Figure 2 shows the considered time series models in our research.

Before we apply these models to analyze Apple's keywords, we check whether the keywords time series data have autoregressive conditional heteroscedastic (ARCH) effects. A time series exhibiting conditional heteroscedasticity can be tested by the Engle (1982) ARCH test which is a Lagrange multiplier (ARCH LM) test as shown in Table II. Table II shows diverse statistics and testing results of Apple's main keywords.

Table I.
Four possible pairs of target variable (Y) and covariate (X)

Target variable (Y)	Covariate (X)
User	Interface
Memory	Data
System	Information
System	Present

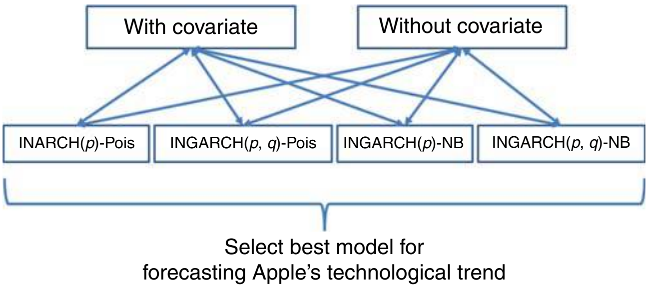

Figure 2.
INARCH procedure for forecasting technological trend

Table II.
Statistics and testing results for Apple's main keywords

Statistic	User	Keyword Memory	System
Mean	147.68	63.1613	183.9677
SD	199.003	76.248	198.605
Skewness	1.46	1.439	0.899
Kurtosis	3.84	3.94	2.59
ARCH LM (1)	20.42 (p-val. = 0.00)	13.73 (p-val. = 0.00)	17.13 (p-val. = 0.00)
ARCH LM (3)	19.92 (p-val. = 0.00)	15.76 (p-val. = 0.001)	15.81 (p-val. = 0.001)
ARCH LM (7)	20.18 (p-val. = 0.005)	14.12 (p-val. = 0.049)	14.73 (p-val. = 0.04)
Ljung-Box (3)	53.35 (p-val. = 0.00)	32.22 (p-val. = 0.00)	42.61 (p-val. = 0.00)
Ljung-Box (7)	62.08 (p-val. = 0.00)	33.12 (p-val. = 0.00)	45.14 (p-val. = 0.00)

In Table II, the mean value of "System" is the largest, this means this keyword is used the most in Apple's technology. The standard deviation of "User" is the biggest of three keywords. That is, the dispersion of this keyword is larger than other keywords. The skewness of "User," "Memory" and "System" are positive. Also the kurtosis for "User" and "Memory" are more than 3, this means high fluctuation in those keywords. The ARCH LM test shows that there exist ARCH effects in main Apple's keywords since the p-values of those main keywords are smaller than 5 percent significant level (0.05). The Ljung-Box test shows that there are serial dependence in those keywords since the p-values of those main keywords are also smaller than 5 percent significant level (0.05). This is a statistical test of independence in the time series data (Box and Pierce, 1970; Ljung and Box, 1978; Harvey, 1993). Therefore we identified the validity to consider the covariate keyword (X) for forecasting target keyword (Y). In the following three figures, main Apple's keywords show a rapidly increasing trend over the years. It means that the time plots of main Apple's keywords are non-stationary even if the autocorrelation function (ACF) plots show one lag autoregressive model may be appropriate, but the partial autocorrelation function (PACF) plots show the moving average model with order 3 or 4. Figure 3 shows the time series plot, ACF, and PACF of "System."

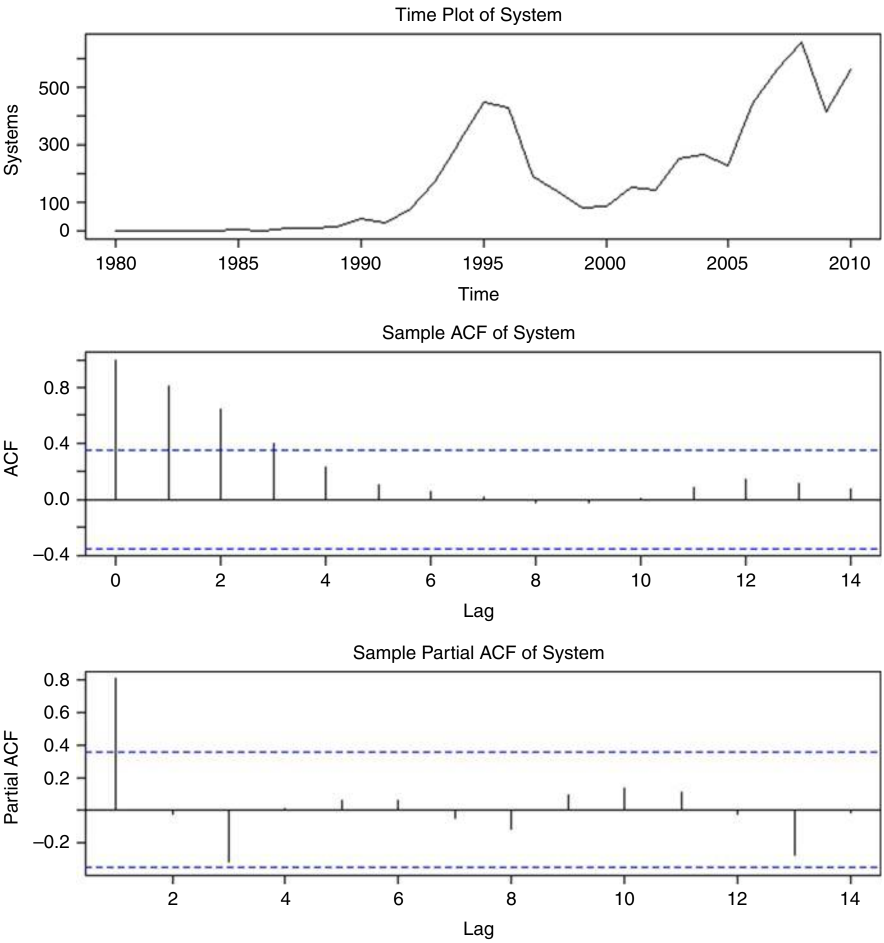

Figure 3.
Time series, ACF, and PACF plots of Apple's keyword: system

From the time series plot in Figure 3, we know that the time series data of keyword "System" are not stationary because they are not contend with the property of sationary that there is no growth and decline (constant through time) in time series data (Bowerman *et al.*, 2005). So we cannot use general and popular time series models such as autoregressive integrated moving average (Bowerman *et al.*, 2005). This is one of the reasons we study INARCH and INGARCH models for the patent keyword analysis. The ACF and PACF remain significantly different from 0 for several time lags. So we can conclude that the time series data of Apple's keyword (System) are not stationary. Figure 4 shows the time series plot and its ACF and PACF.

Like the keyword of "System," the time series data of "Memory" are also not satisfied with the stationary property by the figures of time series plot, ACF, and PACF. Furthermore The PACF of "Memory" represents stronger result of non-stationary than "System." In Figure 5, we can find the non-stationary property of Apple's keyword of "User" by its time series plot, ACF, and PACF. Because of the heterogeneous variance of Apple's keyword data, we should consider ARCH and GARCH structure for fitting non-stationary time series data.

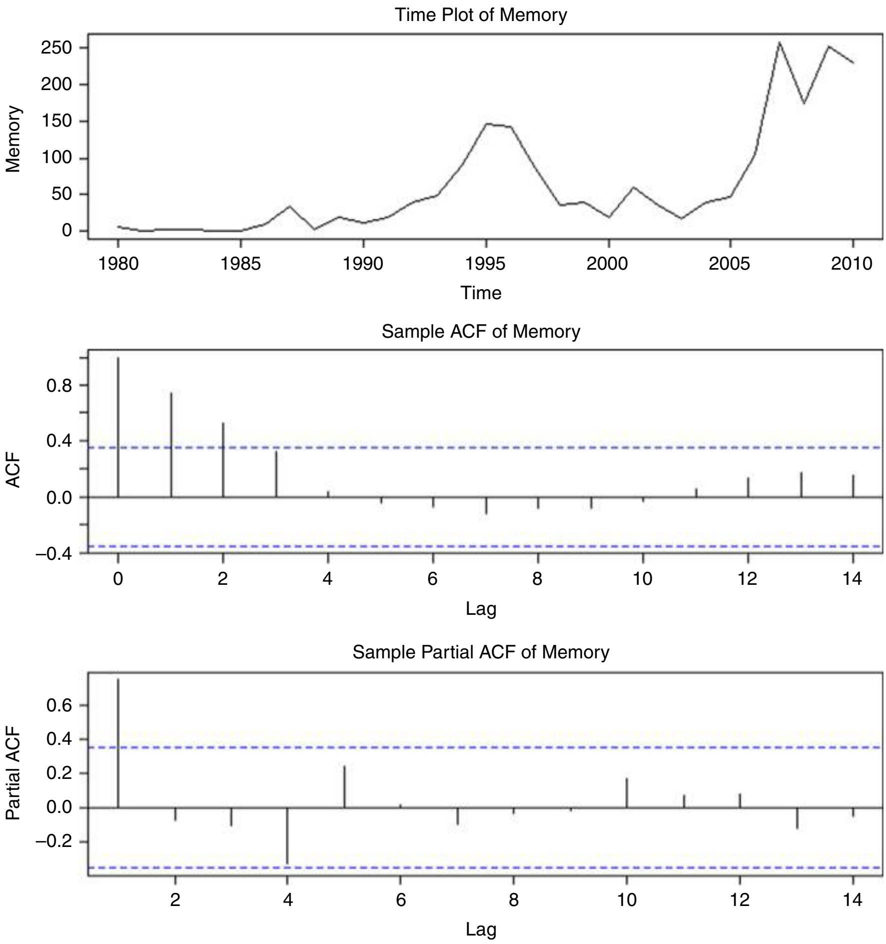

Figure 4.
Time series, ACF, and PACF plots of Apple's keyword: memory

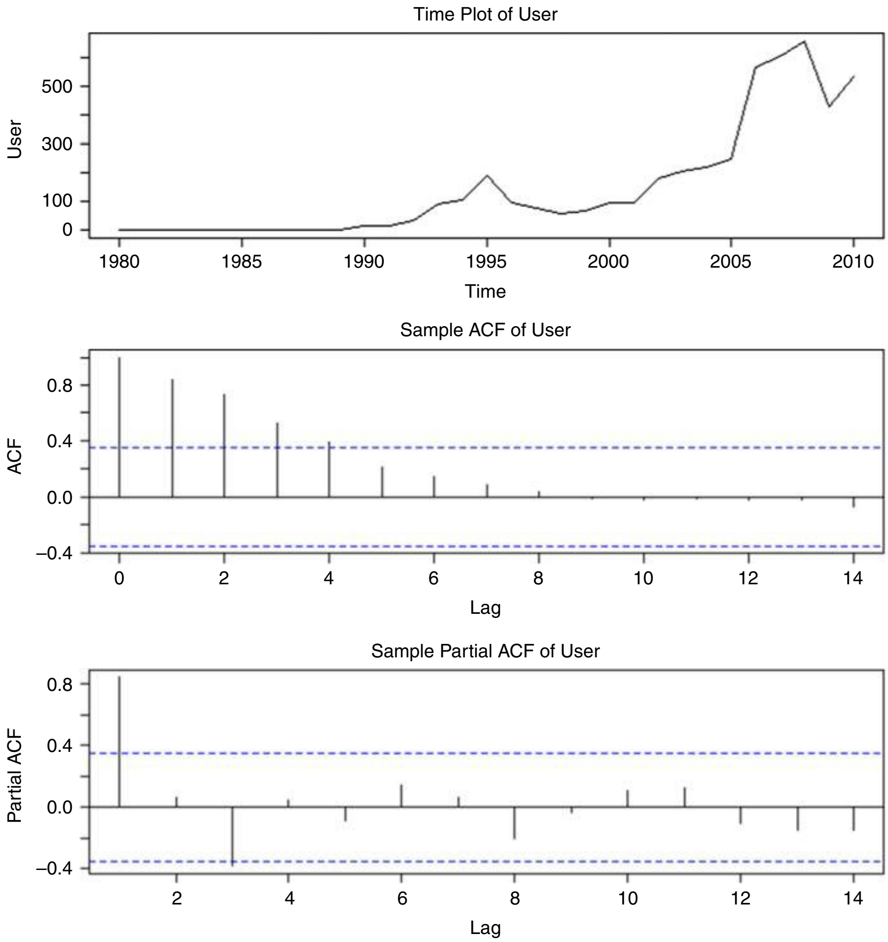

Figure 5.
Time series, ACF, and PACF plots of Apple's keyword: User

Based on the results of Table II and Figures 3-5, the main count time series models are INARCH(1) and INGARCH (1,1) with Poisson and negative binomial distributions. So we applied INARCH(1) and INGARCH (1,1) with Poisson and negative binomial distributions to four possible pairs of target variable (Y) and covariate (X) such as $Y=$User and $X=$Interface, $Y=$Memory and $X=$Data, $Y=$System and $X=$Information, $Y=$System and $X=$Present in Table I. Therefore, we found that the patent keyword data are non-stationary time series data and should be analyzed by count time series models. Table III shows the model comparison of main four Apple's keywords with and without covariates.

From Table III, we can find the fact that INARCH(1)-NB with covariates fitted the Apple's keyword data well than other seven count time series models including INGARCH (1,1)-NB with covariates because the values of AIC and BIC are smallest for the INARCH(1)-NB model among the suggested eight models in this paper. In fact, we did apply the cases in Tables III and IV to the log-linear Poisson autoregression model by Fokianos and Fried (2012), but the values of AIC and BIC by the log-linear Poisson autoregression model are much bigger than the ones by the above suggested eight different count time series models. So we can better predict the trend of the target keyword (User) with the

Table III. Model comparison of main four Apple keywords with covariate

Model	With covariates AIC	With covariates BIC	Without covariates AIC	Without covariates BIC
Y = User and X = Interface				
INARCH(1)	283.206	287.508	1,666.012	1,668.880
INGARCH(1,1)	284.983	290.719	1,479.648	1,483.950
INARCH(1)-NB	*241.789*	*247.525*	313.785	318.087
INGARCH(1,1)-NB	243.486	250.656	323.866	329.602
Y = Memory and X = Data				
INARCH(1)	495.180	499.482	981.903	984.771
INGARCH(1,1)	480.877	486.613	952.065	956.367
INARCH(1)-NB	*277.721*	*283.457*	300.754	305.056
INGARCH(1,1)-NB	281.630	288.800	300.436	306.172
Y = System and X = Information				
INARCH(1)	593.961	598.263	1,883.606	1,886.474
INGARCH(1,1)	595.961	601.697	1,665.584	1,669.886
INARCH(1)-NB	*312.596*	*318.332*	364.232	368.534
INGARCH(1,1)-NB	315.225	322.395	351.738	357.474
Y = System and X = Present				
INARCH(1)	676.619	680.921	1,883.606	1,886.474
INGARCH(1,1)	678.371	684.107	1,665.584	1,669.886
INARCH(1)-NB	*310.883*	*316.619*	364.232	368.534
INGARCH(1,1)-NB	312.623	319.793	351.738	357.474

Table IV. Summary of eight models with and without covariate for Apple's keywords: user and interface

	Y = Target variable and X = covariate						
	With covariates				Without covariates		
Model	α_0	α_1	β_1	γ_1	α_0	α_1	β_1
Y = User and X = Interface							
INARCH(1)-Pois	0.599 (0.285)	0.031 (0.047)	n/a	1.934 (0.087)	5.049 (0.729)	0.987 (0.016)	n/a
INGARCH(1,1)-Pois	0.556 (0.273)	0.000 (0.115)	0.033 (0.102)	1.930 (0.087)	8.2126 (0.966)	0.9694 (0.018)	0.0005 (0.011)
INARCH(1)-NB	0.599 (0.317)	0.031 (0.199)	n/a	1.934 (0.374)	5.049 (1.980)	0.987 (0.200)	n/a
INGARCH(1,1)-NB	0.556 (0.327)	0.000 (0.349)	0.033 (0.323)	1.930 (0.375)	8.2126 (0.966)	0.9694 (0.018)	0.0005 (0.011)
Y = Memory and X = Data							
INARCH(1)-Pois	2.165 (0.645)	0.244 (0.044)	n/a	0.228 (0.013)	10.707 (1.087)	0.893 (0.027)	n/a
INGARCH(1,1)-Pois	1.616 (0.428)	0.017 (0.053)	0.351 (0.077)	0.198 (0.014)	10.400 (1.367)	0.870 (0.037)	0.000 (0.041)
INARCH(1)-NB	2.165 (1.978)	0.244 (0.323)	n/a	0.228 (0.092)	10.707 (4.170)	0.893 (0.227)	n/a
INGARCH(1,1)-NB	1.616 (1.447)	0.017 (0.420)	0.351 (0.585)	0.198 (0.114)	10.400 (6.129)	0.870 (0.279)	0.000 (0.228)

Note: n/a, not available

covariate (Interface) than without the covariate. The same goes for the remaining cases of: Memory, Data; System, Information; and System, Present. Table IV shows the summary of eight models for Apple's keywords (User and Memory) with and without covariates.

Here α_0 represents bias term, and α_1 is the weight on the previous value. In addition, β_1 is the weight on the variance just before one step. Each value in the table consists of estimate and its standard error (value in parentheses). Also γ_1 represent the influence of time series data to variance. The β_1 in ARCH(p) models are n/a (not available) because these models do not consider ω in Section 3.2. In Table IV, INARCH(1) and INGARCH(1,1) using Poisson and negative binomial distributions with covariates show that the estimates of γ_1 are statistically significant. In the INARCH(1)-Pois and INARCH(1)-NB, the estimates of β_1 do not exist in Tables IV and V because the INGARCH(p, q) when $q = 0$ becomes the integer-valued ARCH(p). We also noticed that the covariate in each model contributes in the

Model	α_0	α_1	β_1	γ_1	α_0	α_1	β_1
	Y = Target variable and X = covariate						
	With covariates				Without covariates		
Y = System and X = Information							
INARCH(1)-Pois	2.404 (0.594)	0.279 (0.033)	n/a	1.961 (0.084)	20.478 (1.466)	0.933 (0.015)	n/a
INGARCH(1,1)-Pois	2.400 (0.614)	0.279 (0.052)	0.000 (0.052)	1.960 (0.088)	12.246 (1.325)	0.933 (0.020)	0.011 (0.018)
INARCH(1)-NB	2.404 (1.213)	0.279 (0.326)	n/a	1.961 (0.854)	20.478 (6.029)	0.933 (0.175)	n/a
INGARCH(1,1)-NB	2.400 (1.847)	0.279 (0.546)	0.000 (0.469)	1.960 (0.862)	12.246 (5.395)	0.933 (0.195)	0.011 (0.105)
Y = System and X = Present							
INARCH(1)-Pois	1.264 (0.558)	20.478 (1.466)	n/a	0.933 (0.015)	20.478 (1.466)	0.933 (0.015)	n/a
INGARCH(1,1)-Pois	1.2121 (0.560)	12.246 (1.325)	0.011 (0.018)	0.933 (0.020)	12.246 (1.325)	0.933 (0.020)	0.011 (0.018)
INARCH(1)-NB	1.264 (0.944)	20.478 (6.029)	n/a	0.933 (0.175)	20.478 (6.029)	0.933 (0.175)	n/a
INGARCH(1,1)-NB	1.2121 (1.398)	12.246 (5.395)	0.011 (0.105)	0.933 (0.195)	12.246 (5.395)	0.933 (0.195)	0.011 (0.105)

Note: n/a, not available

Table V. Summary of eight models of system Apple keywords with covariates

model prediction a lot compared to other variables in the models. Similarly, the different models of System of Apple's keyword with covariates in Table V also have the same result as in Table IV.

In particular, the INARCH(1)-NB model of Y = System and X = Present has the smaller AIC than the INARCH(1)-NB model of Y = System and X = Information, as shown in Table III. We can find the reason by looking at the result of Table V. The all estimated coefficients in the INARCH(1)-NB model of Y = System and X = Information are statistically significant rather than in the INARCH(1)-NB model of Y = System and X = Present. In this paper, we found that all γ_1 values of Tables IV and V are significant. This means that the model with covariate are superior to the models without covariate. Figure 6 shows the residual plot of eight different models for Y = User and X = Interface.

We knew that the residuals of the models with covariate are smaller than the models without (W/O) covariate in the Apple's keywords of Y = User and X = Interface. In addition, Figure 6 shows that the residual plots of the INARCH(1)-NB and INGARCH(1)-NB models with covariate have the smaller residuals than six other count time series models. Figure 7 represents the residual plots of all considered time series models in this paper for the keywords of Y = Memory and X = Data.

In Figure 7, we found that the residual ranges of the models with covariate (−50, 50) are smaller than the models without covariate (−100, 100). Also the residuals of the INARCH(1)-NB and INGARCH(1)-NB models with covariate are smaller than other models. Figure 8 shows the residual plots of eight different models for Y = System and X = Information.

The differences between the residual ranges with covariate and without covariate are bigger than the case of the keywords (Y = Memory and X = Data). Like the keywords of "User" and "Memory," Figure 8 shows that the residual plots of the INARCH(1)-NB and INGARCH(1)-NB models with covariate have the smaller residuals than other count time series models. Finally, we show the residual plots of eight different models for Y = System and X = Present in Figure 9.

The residual plots in Figure 9 show the same result as the previous residual plots. In the results from Figures 6 to 9, we found that the residuals with covariates were smaller than without covariates. Next we want to do technology forecasting with the count time series models in this paper.

5. Forecast accuracy measures for Apple's keywords

In this section, we want to find the pairs which have high correlation with each other. Using these pairs with significantly high correlation, we can forecast the trends of the target keywords representing Apple's technologies. Table VI shows the residuals of INARCH(1)-NB

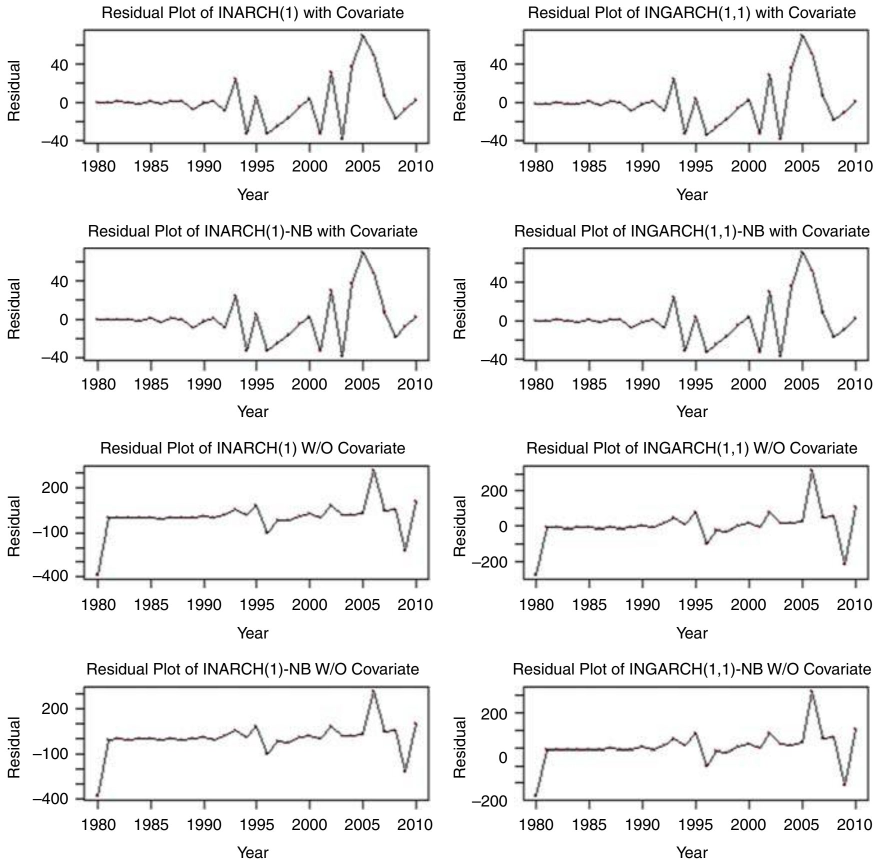

Figure 6.
Residual plots of eight different models for $Y =$ User and $X =$ Interface

model with single covariate over year and evaluation results by mean absolute error (MAE) and root mean squared error (RMSE).

In Table VI, we used the two most popular measures of forecasting as MAE and RMSE. Let Y_i denote the ith observation in time series data and $\hat{Y}_i$ denote forecasted value of Y_i. The forecast error is simply $e_i = Y_i - \hat{Y}_i, (i = 1, 2, \ldots, n)$ which is on the same scale as the data. The two most commonly used scale-dependent measures are based on the absolute errors or squared errors (Bowerman *et al.*, 2005). We show the MAE and RMSE as follows:

$$\text{Mean absolute error(MAE)} = \frac{\sum_{i=1}^{n} |e_i|}{n} \tag{13}$$

$$\text{Root mean squared error(RMSE)} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n}} \tag{14}$$

The MAE measure is the average of the absolute deviation for all-time series data. Also the RMSE is a measure of the average of the root mean squared errors for all observations and forecasted values. We found the Apple keywords pairs (Y, X) so that the INARCH(1) model

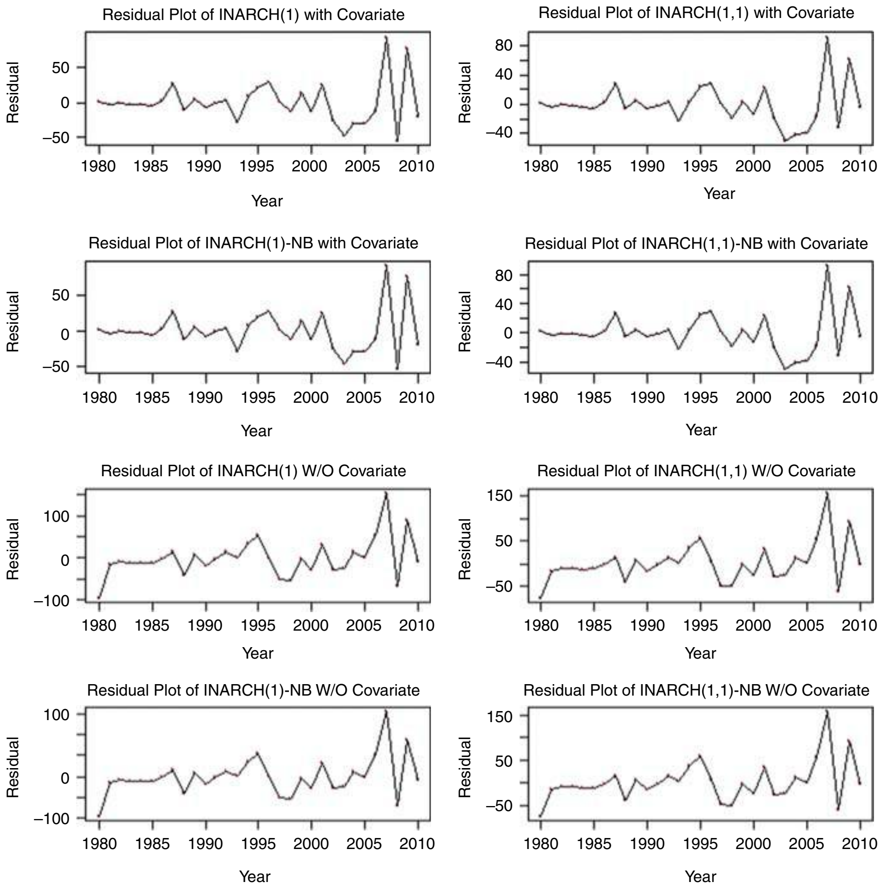

Figure 7.
Residual plots of eight different models for Y = Memory and X = Data

which was performed for model fitting in the previous session was applied as shown in Table VI. The results of the INARCH(1) model were surprisingly good. For example, the both results when Y = Information and X = Data and when Y = Data and X = Information have significantly good predictions. In Table VI, we can find these similar results easily. Based on these results, we can forecast Technology in the near future. In addition, we knew that the residual value of a keyword grows according to its increasing frequency. In (Y, X) of Table VI, the yearly change of residual means the change of frequency over year. Figure 10 supports the high correlation of Apple's keywords pairs in Table VI.

The horizontal and vertical axes of Figure 10 represent time period and keyword frequency, respectively. For example, the correlation coefficient between "device" and "electronic" is very large and we find that the time series plots of "device" and "electronic" are very similar in Figure 10. Most keywords have first peak near the 1995. At this period, Apple applied the researched and developed technologies to patents and dominated to the personal computer market by the product of "Power Macintosh." Since then Apple lost the market to IBM and Microsoft. After the hard times of late 1990s through early 2000s, Apple has developed dramatically based on the innovative products such as iPod, iPhone, and iPad. So in Figure 10, the trends of most keywords have increased since late 2000s. Especially the keywords of "device," "electronic," "information," "image," "power," "control,"

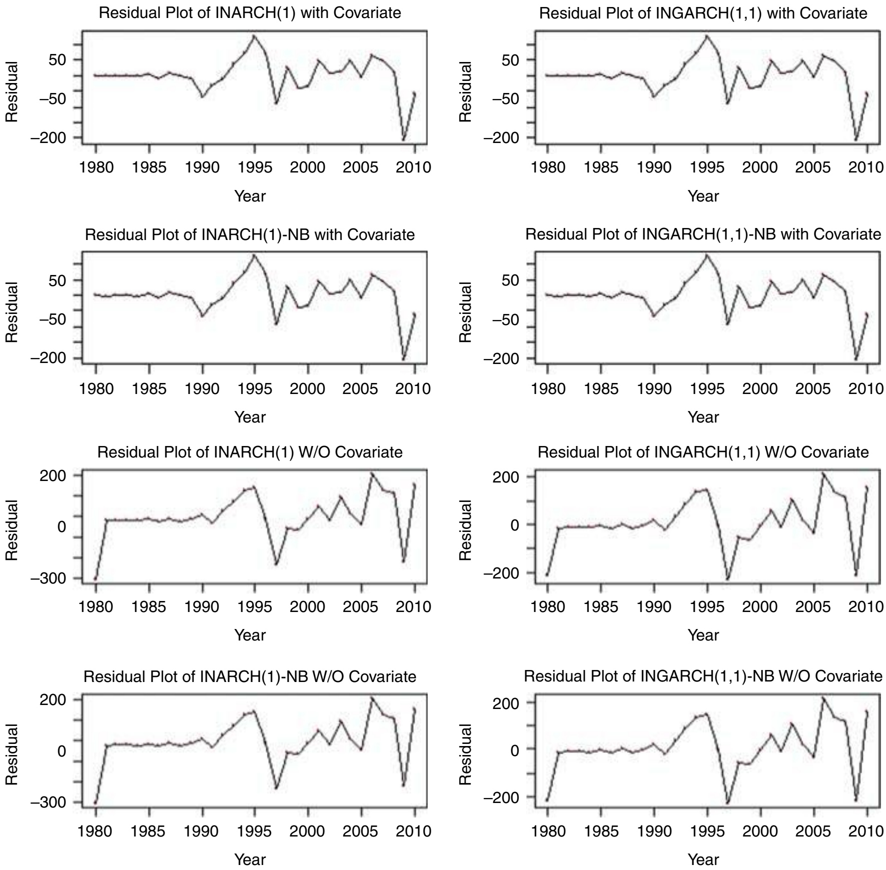

Figure 8.
Residual plots of eight different models for $Y =$ System and $X =$ Information

"audio," "portable," "circuit," "time" have been increased rapidly. So we can understand that the importance of the technologies based on these keywords have grown for Apple. Therefore, we show the result of Apple's technology analysis by the proposed methodology in Figure 11.

The target technology of Apple is based on the keywords of "User," "Memory," and "System." The keywords of "Interface," "Data," "Information," and "Present" are covariate with the target keywords. In the technological structure, each keyword represents an assigned technology. In addition, the technological keywords of "Device," "Electronic," "Image," "Power," "Control," "Audio," "Portable," "Circuit," and "Time" are needed to be considered to the target keywords. So, the technologies of Apple are based on the technological structure in Figure 11. Using the relations between Apple's technologies, Apple can perform his research and development (R&D) planning for technological innovation and new product development. Also we can forecast technological trends of Apple by the IN-ARCH(1) time series model.

6. Conclusions

We proposed integer-valued GARCH processes for Apple's technology analysis based on time series keyword data from retrieved patent documents. Most previous researches for

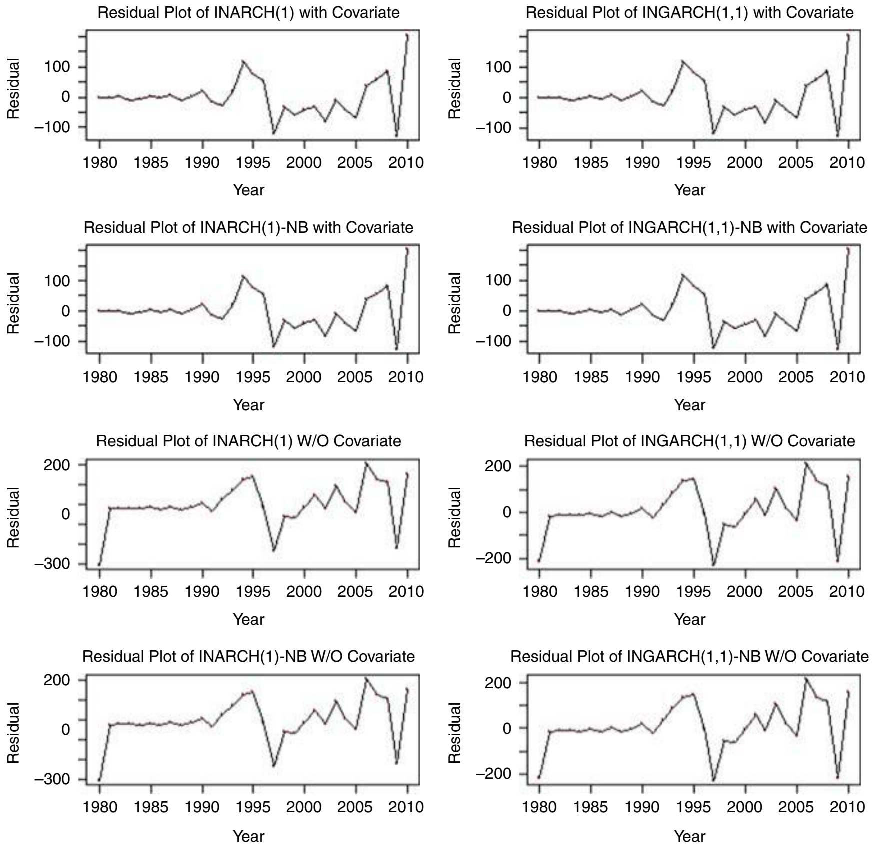

Figure 9.
Residual plots of eight different models for Y = System and X = Present

time series patent data used continuous time series models. But the technological keyword data from patent documents are not continuous, they are count data. So we used count time series models such as integer-values time series models such as GARCH and ARCH based on Poisson and negative binomial distribution. Also we considered two correlated keywords at the same time instead of one single keyword, so we constructed the time series models with covariates. From this study, we found that how much the target response keyword could be predicted by the predictor keyword by count time series modeling. That is, technologies of predictor keywords could affect the technological developments of target response keywords. In this research, we found the best time series model for predicting trends of Apple's technological keywords is IN-ARCH(1). Our selected model contributes to Apple's domain experts for planning Apple's R&D strategy. In this paper, we considered the technologies related to "user," "memory," and "system" for target variables, and the technologies of "interface," "data," "information," and "present" were used for covariate variables with target variables. From the results of our experiments, we found the technological trends of Apple's technologies by the count time series models. The GARCH models with covariates are superior to the models without covariates. This means that the various technologies of Apple have been developed by influencing each other. In real R&D planning of Apple, the time series modeling of a technology should be constructed by considering other technologies (keywords).

Table VI. Residual, MAE, and RMSE of INARCH(1)-NB model with single covariate

(Y,X)	2007	2008	2009	2010	MAE	RMSE
(device, electronic)	−37.565	−387.264	−536.317	−360.675	330.456	377.196
(data, information)	23.484	−9.657	−117.747	69.697	55.146	69.582
(system, network)	45.860	24.174	−45.072	−34.515	37.405	38.439
(user, display)	−103.703	−24.909	−238.072	−68.571	108.814	134.866
(media, user)	−51.091	139.019	−360.838	−172.065	180.753	213.159
(display, data)	204.810	0.819	−13.806	121.266	85.175	119.210
(computer, invention)	−53.613	−42.829	−64.865	−0.880	40.546	47.214
(electronic, device)	103.969	339.374	25.948	171.230	160.130	197.470
(interface, display)	−31.417	34.887	−66.043	−41.025	43.343	45.412
(information, data)	6.113	14.840	17.355	−5.776	11.021	12.167
(image, power)	−54.804	−67.130	−128.933	−14.393	66.315	78.008
(memory, information)	117.799	−45.652	76.901	7.105	61.864	74.035
(power, device)	−97.900	−189.870	5.657	−167.872	115.325	135.875
(control, data)	10.828	45.342	67.939	56.687	45.199	50.006
(signal, data)	33.576	−34.572	12.295	−1.024	20.367	24.874
(video, device)	−10.671	−136.330	−126.267	−135.579	102.212	115.136
(plurality, display)	−22.830	−28.037	−31.896	−70.383	38.287	42.657
(audio, power)	−61.289	43.226	108.695	−185.964	99.794	114.042
(portable, device)	−125.213	−41.058	−80.478	−129.897	94.162	100.889
(application, information)	47.161	95.938	−34.880	44.506	55.621	60.468
(invention, system)	−25.609	−37.917	−32.967	−73.345	42.459	46.260
(circuit, power)	84.407	26.488	52.822	−44.649	52.091	56.147
(present, system)	26.658	−26.385	−14.233	−26.806	23.520	24.124
(set, data)	−28.099	2.462	−3.590	13.277	11.857	15.691
(portion, electronic)	−26.978	−77.524	−252.339	−43.450	100.073	134.444
(digital, set)	62.307	−56.333	35.532	2.848	39.255	45.624
(content, interface)	169.123	129.752	51.322	144.130	123.582	131.192
(object, system)	14.978	−2.663	−17.177	7.586	10.601	12.083
(color, object)	14.090	−28.453	−4.315	−11.598	14.614	17.038
(time, display)	−23.323	−27.332	−20.119	−52.991	30.941	33.555
(network, system)	−5.036	−0.872	−13.345	19.130	9.595	11.939

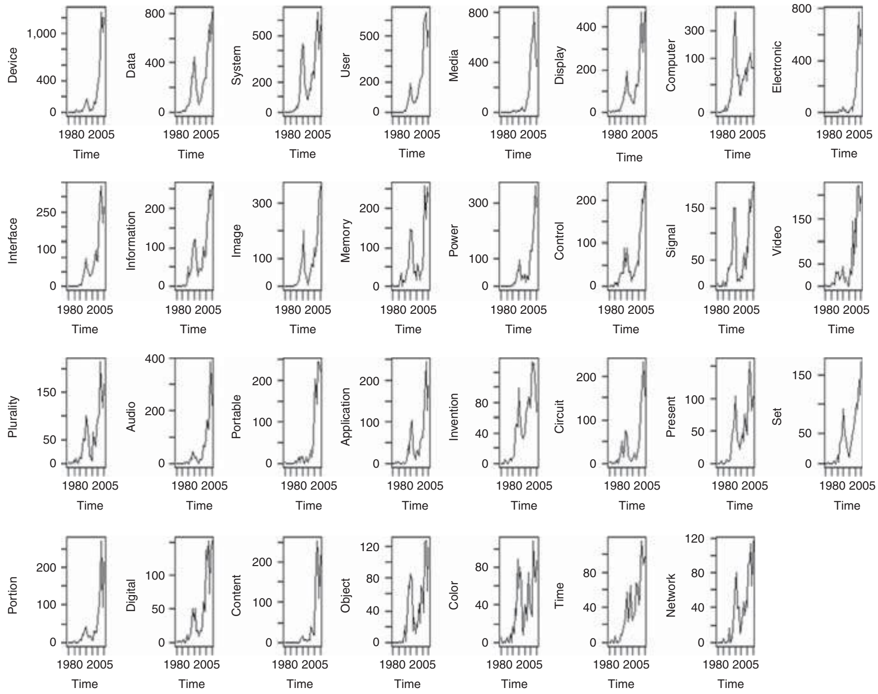

Figure 10.
Time plots of 31 Apple keywords

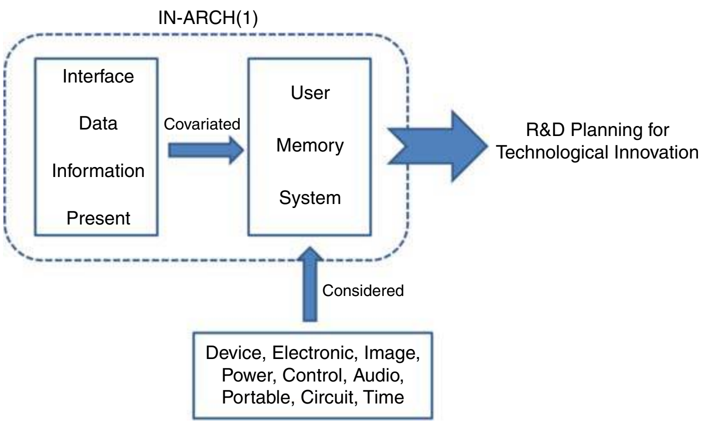

Figure 11.
Result of Apple's technology analysis

Furthermore, we recommend IN-ARCH(1) for the forecasting model of Apple technology. The proposed methodology could be applied to diverse technology companies such as Samsung, Google, etc. for understanding their technological trends between technological keywords. In this paper, we applied the methodology to only one company, but our approach is expected more performance to analyze the technological trends between the competitors which are Apple and Samsung. In our future study, we will compare the technological trends of many competitive companies for more advanced integer-valued time series models.

References

Akritas, M. (2016), *Probability & Statistics with R for Engineers and Scientists*, Pearson, Boston, MA.

Arruda-Filho, E.J. and Lennon, M.M. (2011), "How iPhone innovators changed their consumption in iDay2: hedonic post or brand devotion", *International Journal of Information Management*, Vol. 31 No. 6, pp. 524-532.

Arruda-Filho, E.J., Cabusas, J.A. and Dholakia, N. (2010), "Social behavior and brand devotion among iPhone innovators", *International Journal of Information Management*, Vol. 30 No. 6, pp. 475-480.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, Vol. 31 No. 3, pp. 307-328.

Bowerman, B.L., O'Connell, R.T. and Koehler, A.B. (2005), *Forecasting, Time Series, and Regression, An Applied Approach*, Brooks/Cole, Independence, KY.

Box, G.E.P. and Pierce, D.A. (1970), "Distribution of residual correlations in autoregressive-integrated moving average time series models", *Journal of the American Statistical Association*, Vol. 65 No. 332, pp. 1509-1526.

Christou, V. and Fokianos, K. (2014), "Quasi-likelihood inference for negative binomial time series models", *Journal of Time Series Analysis*, Vol. 35 No. 1, pp. 55-78.

Davis, R.A. and Wu, R. (2009), "A negative binomial model for time series of counts", *Biometrika*, Vol. 96 No. 3, pp. 735-749.

Engle, R.F. (1982), "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation", *Econometrica*, Vol. 50 No. 4, pp. 987-1007.

Engle, R.F. and Ng, V.K. (1993), "Measuring and testing the impact of news on volitality", *Journal of Finance*, Vol. 48 No. 5, pp. 1749-1778.

Feinerer, I. and Hornik, K. (2014), *Package 'tm' Ver. 0.6, Text Mining Package*, CRAN of R Project, Vienna.

Ferland, R., Latour, A. and Oraichi, D. (2006), "Integer-valued GARCH process", *Journal of Time Series Analysis*, Vol. 27 No. 6, pp. 923-942.

Fokianos, K. and Fried, R. (2010), "Interventions in INGARCH processes", *Journal of Time Series Analysis*, Vol. 31 No. 3, pp. 210-225.

Fokianos, K. and Fried, R. (2012), "Interventions in log-linear Poisson autoregression", *Statistical Modelling*, Vol. 12 No. 4, pp. 299-322.

Funk, J.L. (2011), "Standards, critical mass, and the formation of complex industries: a case study of the mobile Internet", *Journal of Engineering and Technology Management*, Vol. 28 No. 4, pp. 232-248.

Guidolin, M. and Guseo, R. (2014), "Modelling seasonality in innovation diffusion", *Technological Forecasting and Social Change*, Vol. 86, pp. 33-40.

Halal, W.E. (2013), "Forecasting the technology revolution: results and learnings from the TechCast Project", *Technological Forecasting and Social Change*, Vol. 80 No. 8, pp. 1635-1643.

Harvey, A.C. (1993), *Time Series Models*, 2nd ed., Harvester Wheatsheaf, New York, NY.

Hong, J., Shin, J. and Lee, D. (2016), "Strategic management of next-generation connected life: focusing on smart key and car-home connectivity", *Technological Forecasting and Social Change*, Vol. 103, pp. 11-20.

Hung, C.-Y., Lee, W.-Y. and Wang, D.-S. (2013), "Strategic foresight using a modified Delphi with end-user participation: a case study of the iPad's impact on Taiwan's PC ecosystem", *Technological Forecasting and Social Change*, Vol. 80 No. 3, pp. 485-497.

Hunt, D., Nguyen, L. and Rodgers, M. (2007), *Patent Searching Tools & Techniques*, Wiley, Hoboken, NJ.

Joseph, M.H. (2011), *Negative Binomial Regression*, 2nd ed., Cambridge University Press, Cambridge.

Jun, S. (2013), "A new patent analysis using association rule mining and Box-Jenkins modeling for technology forecasting", *Information-An International Interdisciplinary Journal*, Vol. 16 No. 1B, pp. 555-562.

Jun, S. and Park, S. (2013), "Examining technological innovation of Apple using patent analysis", *Industrial Management & Data Systems*, Vol. 113 No. 6, pp. 890-907.

Jun, S. and Uhm, D. (2010), "Technology forecasting using frequency time series model: bio-technology patent analysis", *Journal of Modern Mathematics and Statistics*, Vol. 4 No. 3, pp. 101-104.

Jung, R.C., Kukukb, M. and Liesenfeldc, R. (2006), "Time series of count data: modeling, estimation and diagnostics", *Computational Statistics & Data Analysis*, Vol. 51 No. 4, pp. 2350-2364.

Kim, J.-M. and Jun, S. (2015), "Graphical causal inference and copula regression model for Apple keywords by text mining", *Advanced Engineering Informatics*, Vol. 29 No. 4, pp. 918-929.

Lakka, S., Michalakelis, C., Varoutas, D. and Martakos, D. (2013), "Competitive dynamics in the operating systems market: modeling and policy implications", *Technological Forecasting and Social Change*, Vol. 80 No. 1, pp. 88-105.

Liboschik, T., Fried, R., Fokianos, K., Probst, P. and Rathjens, J. (2015), Package 'tscount' Ver. 1.4.0, Analysis of Count Time Series, CRAN of R Project, Vienna.

Ljung, G.M. and Box, G.E.P. (1978), "On a measure of lack of fit in time series models", *Biometrika*, Vol. 65 No. 2, pp. 297-303.

Nam, S., Nam, C. and Kim, S. (2015), "The impact of patent litigation on shareholder value in the smartphone industry", *Technological Forecasting and Social Change*, Vol. 95, pp. 182-190.

Park, S. and Jun, S. (2012), "New technology management using time series regression and clustering", *International Journal of Software Engineering and Its Applications*, Vol. 6 No. 2, pp. 155-160.

Park, S., Kim, J., Jang, D., Lee, H. and Jun, S. (2016), "Methodology of technological evolution for three-dimensional printing", *Industrial Management & Data Systems*, Vol. 116 No. 1, pp. 122-146.

R Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, available at: http://www.R-project.org/ (accessed June 1, 2014).

USPTO (2015), "The United States Patent and Trademark Office", available at: www.uspto.gov/ (accessed January 1, 2015).

West, J. and Mace, M. (2010), "Browsing as the killer app: explaining the rapid success of Apple's iPhone", *Telecommunications Policy*, Vol. 34 No. 5, pp. 270-286.

WIPSON (2014), "WIPS Corporation", available at: www.wipson.com/ (accessed July 24, 2014).

Wonglimpiyarat, J. (2005), "Standard competition: is collaborative strategy necessary in shaping the smart card market?", *Technological Forecasting and Social Change*, Vol. 72 No. 8, pp. 1001-1010.

Zeger, S.L. (1998), "A regression model for time series of counts", *Biometrika*, Vol. 75 No. 4, pp. 621-629.

Zhu, F. (2011), "A negative binomial integer-valued GARCH model", *Journal of Time Series Analysis*, Vol. 32 No. 1, pp. 54-67.

Further reading

Fagerberg, J. (1987), "A technology gap approach to why growth rates differ", *Research Policy*, Vol. 16 No. 2, pp. 87-99.

Hall, B.H., Jaffe, A.B. and Trajtenberg, M. (2001), "The NBER patent citation data file: lessons, insights and methodological tools (No. w8498)", National Bureau of Economic Research, Cambridge, MA.

Kim, H. and Song, J. (2013), "Social network analysis of patent infringement lawsuits", *Technological Forecasting and Social Change*, Vol. 80 No. 5, pp. 944-955.

Liboschik, T., Kerschke, P., Fokianos, K. and Fried, R. (2014), "Modelling interventions in INGARCH processes", *International Journal of Computer Mathematics*, Vol. 93 No. 4, pp. 1-18.

Van de Kaa, G. and De Vries, H.J. (2015), "Factors for winning format battles: a comparative case study", *Technological Forecasting and Social Change*, Vol. 91, pp. 222-235.

Corresponding author
Sunghae Jun can be contacted at: statcs@gmail.com