Paper

Analysis and Modeling of Domain Registration Process

Piotr Arabas[a,b], Przemysław Jaskóła[b], Mariusz Kamola[a,b], and Michał Karpowicz[b]

[a] *Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw, Poland*
[b] *Research Academic Computer Network (NASK), Warsaw, Poland*

Abstract—The paper presents analysis of the domain name reservation process for the polish `.pl` domain. Two models of various time scale are constructed and finally combined to build long range high resolution model. The results of prediction are verified using real data.

Keywords—domain market, domain registration, forecasting, time series modeling.

1. Introduction

In time of rapid growth of Internet, domain names became an important commodity [1]. In consequence, the volume of DNS market became dependent on overall economic conditions and expectedly follows standard laws of demand, and supply. Furthermore, as the number of attractive domain names is limited, there exists possibility of investing and earning relatively high profits. For all these reasons domain registration statistics present interesting set of data to be analyzed. The aim of this article is to present results of analysis and modeling of domain registration process.

Similar analysis were presented in [2], [3], [4], while the secondary market was studied in [5], however none of these papers covered Poland. Much broader literature is devoted to semantic analysis of domain names, which can be used to assess their qualities [6], [7] or pricing [8], [9]. As far as some of the results of these works have direct connection with demand modeling and pricing domain names, they, in our opinion, neglect the most basic behavior of domain users.

In this paper, we concentrate on primary market (registration) modeling. We try to find out some specific characteristics of this process using abundant data of Polish domain registry. First, we try to identify its general properties by analyzing basic statistics in various time scales and applying harmonic analysis to determine characteristic periods. We show that data conform to some patterns, two of them – weekly and yearly – being most obvious. Following this observation, we propose to construct specialized models on both time scales and, possibly, compose more complex models of them. It must be noted that even short horizon modeling may provide valuable predictions, e.g., for planning of an advertising campaign.

The rest of the paper is organized as follows: in Section 2 we describe the problem, which is subject of this research. Next, in the Section 3 we show related work and draw our solution. In the Section 4 the data and basic characteristics of the process are presented along with results of a preliminary analysis. The Section 5 presents the model built to reflect long-range behavior of registration process together with the results of one-year ahead prediction. The short range model and results of its verification are described in the Section 6. Then, in Section 7 we combine both models into a composite model allowing one year prediction with resolution of one day. We conclude in Section 8.

2. Description of the Problem

The domain names are organized in a hierarchical manner, with the last part of each name being a name of top level domain (TLD). Important portion of TLDs are national domains with `.pl` being polish TLD. The registry of each TLD is kept by some institution designated by ICAN, being responsible for domains worldwide. In Poland, such registry for `.pl` domain, together with various regional, functional etc. sub-domains is managed by NASK (Research and Academic Computer Network). The interest in analyzing and modeling of the domain registration process is caused by several factors. First of all, registration is a commercial activity with fees paid for registration and then, repetitively, each year for prolonging domain activity. NASK sells domains mostly on the wholesale market to the number of companies offering various other network services to end users. It must be noted that domains are not only bought by companies or individuals who need to establish a new internet service, e.g., webpage, but also (as mentioned earlier) as a kind of investment, for future resale on the secondary market.

The result of this segmentation are different behaviors of various groups of clients – big companies are possibly less price sensitive than individual users, however, most sensitive and in the fact chimeric group may be the investors. This group may also have different strategies of renewing domains – some domains which are not needed (e.g., then turned out to be unprofitable) may be dropped and some may be re-registered after short time. Although we do not analyze renewal of domains here and neglect influence of its price on registration process it is important to realize that periodic expiration of a large number of domains may result in apparently spontaneous accumulation of re-registrations.

3. Proposed Solution

As it was mentioned in the introduction, the body of work related to modeling of domain registration process is relatively scarce left aside papers devoted to semantic analysis of domain names. What we try to do is to analyze of the registration process as a whole – we do not distinguish more and less valuable domains, as NASK sells them on the wholesale basis without such differentiation. We also decided not to model price factor to simplify the model. In a fact, we tried modeling price–demand correlation using some basic economic models, e.g., Cobb-Douglas or Gutenberg [10], however we found it ineffective and possible unnecessary. The reason was relatively scarce amount of data resulting from rare and usually too small changes in the pricing strategy. In the analyzed period, only one price change had clearly visible effect – it was lowering the registration price in 2008 (see Fig. 1). Furthermore, our aim was to construct models that could be used for prediction on some clearly defined horizon, which application allows considering external factors as constant and recalculating models if necessary.

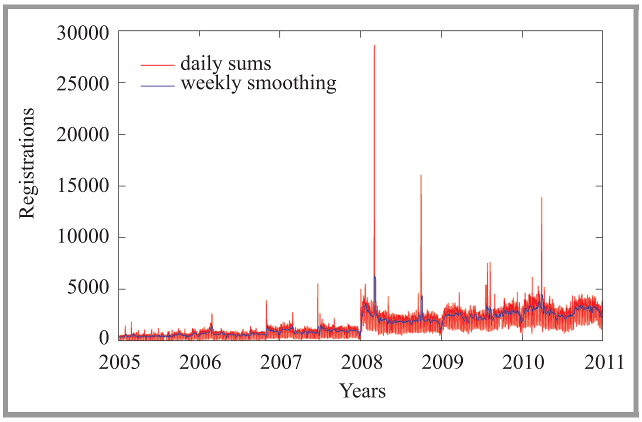

Fig. 1. Domain registrations before removing anomalies.

With such assumptions the process may be modeled as pure time series, which allows the use of well known methodology (see, e.g., [11], [12]). The approach is well grounded for modeling economic and sociology data, and we suppose that our case does not differ much from, e.g., air travel frequency [13], or real estate prices [14]. The basic assumption, which we adopted after, e.g., [12], [15] is that the base process (domain registration in this case) may be decomposed in the following way:

$$x_t = p_t + s_t + e_t, \tag{1}$$

where p_t is trend, s_t is the seasonal and e_t is the irregular component. The approach is natural since trend can be easily observed in the registration data (see Fig. 1), it will be also shown in the next section that the seasonal component is even stronger.

In economic modeling the seasonality is typically defined as periodic process corresponding to yearly cycle (see, e.g., [15], [16]), however the same technique may be used to other, longer or shorter periods. In a fact, it is typical for many processes to exhibit seasonality on several timescales, the best example being presence of short and long economy cycles (waves) [12], [17], or even infinite number of time scales like for self-similar processes [18].

The models used to describe seasonality range from relatively simple periodic (e.g., trigonometric) functions to complex formulas involving regression and relying on expert knowledge, some of them being recognized standards, like X-12 or STL [12], [15], [19]. Other techniques incorporate some approximation methods like, e.g., wavelet analysis [20]. Although using such complex models allows attain precision and draw from rich experience of other researchers, we limited our work to application of the simplest models based on calculation of seasonal means [12], while we tried exploring various time-scales of analyzed process, and finally constructing a model covering all time scales. We did it for two reasons: first, the results of such modeling are simpler to interpret so it is possible to assess the most important properties of registration process clearly. Next, as the aim of the work was prediction, it is easier to build stable forecasts using simpler (i.e., having less parameters, but also needing less restricting assumptions) models.

4. Data

Data were made available by Polish domain registry and consisted of daily sums of registered domains in years 2005–2010. All kinds of domains in polish `.pl` domain, i.e., regional, functional, etc. were summed up. The data were in raw format, as directly dumped from system logs and contained some irregularities. There were two sorts of them:

- missing or duplicated samples of extremely low value,
- samples of anomalously high value.

The first group may be associated with malfunction of the infrastructure, mainly the database software. The second kind of anomalies is mainly caused by some extraordinary promotions, resulting in higher than usual sales; it can be easily observed in the Fig. 1. Fortunately, there were only two gaps in data, which we decided to interpolate. Also, some additional data cleaning had to be performed.

Anomalously high values pose much more problems, as we cannot precisely isolate them by analyzing registration history only. Another important question is what value should be inserted instead of anomalous sample. We decided to be very conservative and deal only with these samples, which we can associate with known marketing campaigns. With help of marketing division staff we identified two such events in 2008, and another two in 2009. Furthermore, we were able to assess number of domains registered during these campaigns, which in turn allowed us to subtract them from appropriate samples. We did not eliminate one possible anomaly in the beginning of 2010, as we could not identify its cause. The data after cleaning are depicted in Fig. 2.

The filtered data contain some likely anomalies still, however they are not so high like those removed, and do not

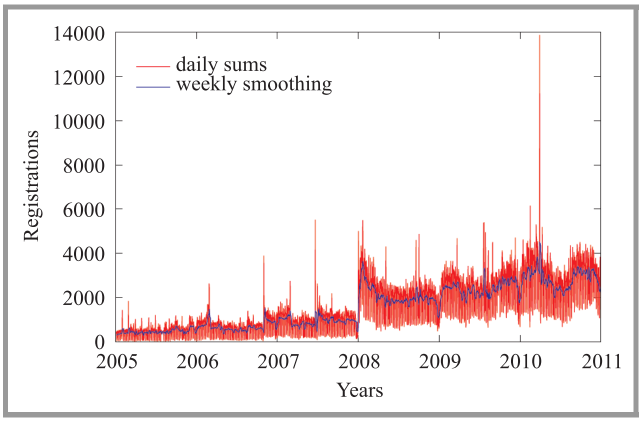

Fig. 2. Domain registrations after removing identified anomalies.

influence smoothed data visibly (except mentioned earlier and left unfiltered anomaly in 2010). Thanks to this it is possible to make some important observations – first of all, the number of registered domains grows, the trend is however disturbed by one rapid rise in the beginning of year 2008. The phenomenon may be easily explained by significant lowering of registration fee in that year. It must be noted that after a change in pricing strategy in 2008, registration price was much lower than renewal fee. In the result, many domains, which were probably bought as a kind of investment, are dropped after one year, while another are re-registered in the beginning of next year, and give cause to some rise in first months of each year.

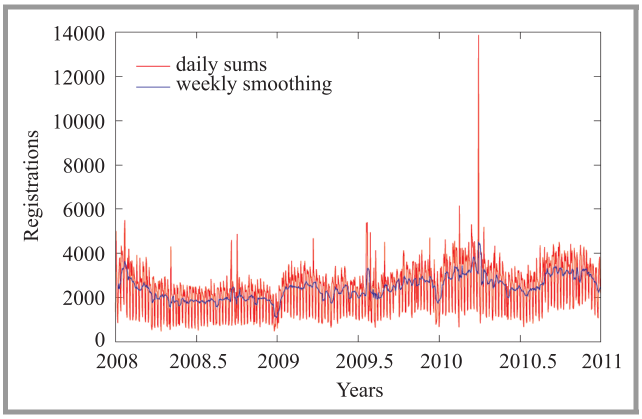

Fig. 3. Domain registrations for last three years after removing identified anomalies.

To observe yearly changes it is better to have a look at graph presenting only 3 years (Fig. 3). The data show visible yearly pattern – manifesting mainly in very low number of registrations during winter holidays and also some higher frequency variations, which may be easily identified as weekly cycles. To emphasize these variations, another set of graphs depicting each year separately is presented in Fig. 4. Yearly patterns may be observed in monthly aggregated data presented in the analogous set of diagrams – see Fig. 5. Summing information from both set of graphs, it must be said that weekly pattern is clearly visible and

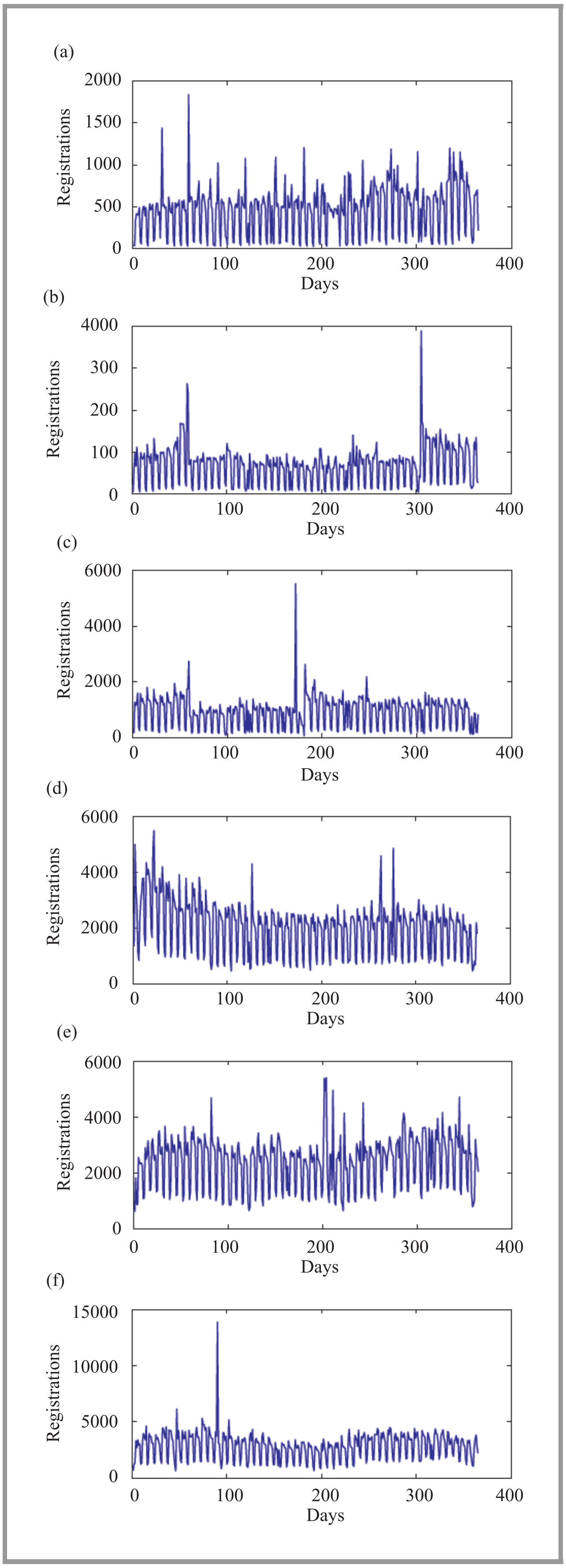

Fig. 4. Registrations in years: (a) 2005; (b) 2006; (c) 2007; (d) 2008; (e) 2009; (f) 2010.

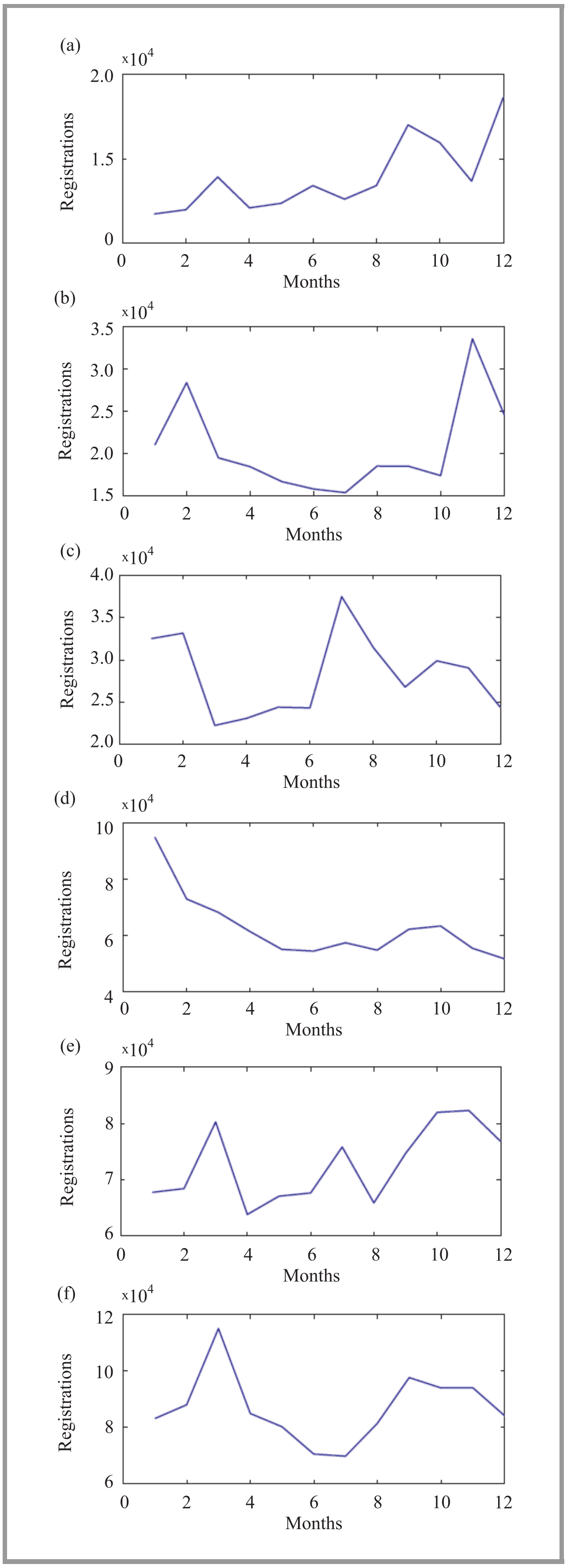

Fig. 5. Monthly sums of registrations in years: (a) 2005; (b) 2006; (c) 2007; (d) 2008; (e) 2009; (f) 2010.

relatively regular, while yearly pattern has rather vague character. It is possible to identify two periods of lower sales during the year – first, more noticeable and easier to locate is winter holidays. The second could be associated with summer holidays, however it tends to move around.

To check for existence of other characteristic periods, we applied spectral analysis by computing power spectrum for period 2005–2009 (see Fig. 6). In this case, we skipped last year as it is used for verification of models presented in the next sections. The number of analyzed samples is too small to gain significant results for longer periods (e.g., one year), however period of one week is again clearly visible. Another period equals approximately to half a week may be treated as a kind of harmonic frequency, and can be explained by the shape of weekly pattern.

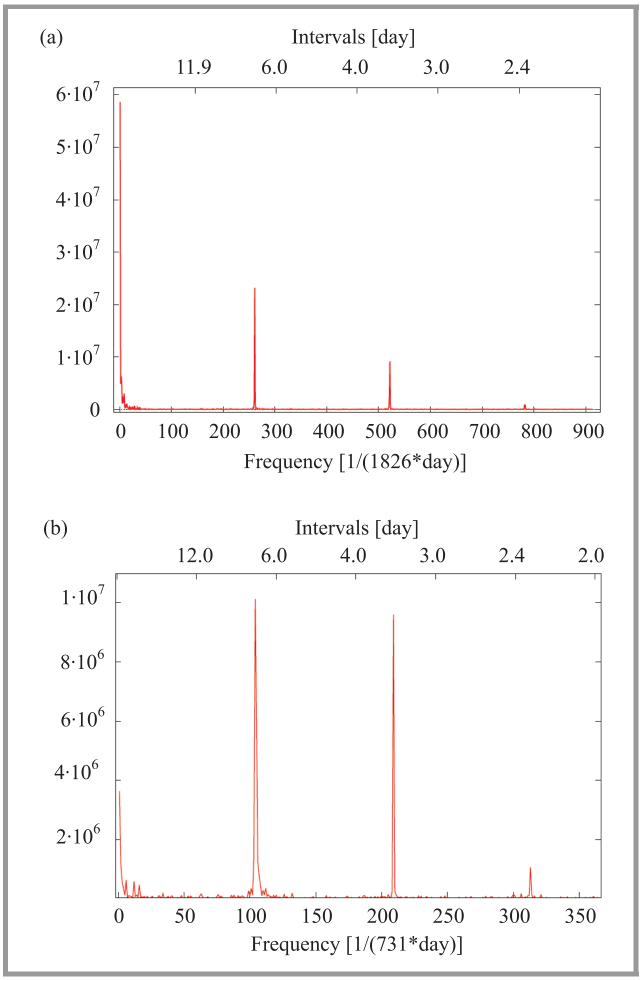

Fig. 6. Power spectrum of the registration process in years 2005–2009.

The most important result of these preliminary analysis is identification of two characteristic periods of the registration process: shorter with length of one week and longer associated with yearly variations. Following these observations, we decided to build two models describing longer and shorter cycles separately to simplify their construction and to allow further analysis.

5. Long Range Modeling

When making strategic decisions like setting new prices or planning a capacity of DNS servers, it is useful to have an estimation of the future sales. Such a prediction can be built upon appropriately designed model and, the most important requirement is a prediction horizon long enough – at least one year. On the other hand, there is no need for high temporal resolution – predicting sales in subsequent months is typically sufficient. After initial analysis we decided not to model influence of domain prices on registrations. There were two reasons for this: first the price changes are relatively rare so it is difficult to gather data necessary to identify any model. But the situation is even more complex, as end users do not observe NASK prices being wholesale prices for dealers. Every dealer has his/her own pricing strategy, furthermore domain names are often sold as a part of a bundle – together with Internet access, web service or mailbox.

5.1. Seasonality and Trend

For the above reasons, we decided to treat the registration process as a time series and build a model using the most classical approach, i.e., to estimate the trend and seasonality first. Then, having as we hoped stationary residuals, we planned to fit an autoregressive process to them. For identification we used monthly aggregated data from period of 2005–2009, and then 2008–2009, while we used data from 2010 for verification.

Such shortening of the learning period is the result of a rapid jump in registrations after lowering prices in 2008, what can be best seen in the graph in Fig. 7 showing two trends fitted to deseasonalized data. Values for the last twelve months in the graph Fig. 7 are predictions for year 2010 – it can be easily seen that including rise in 2008 in unfiltered form results in excessive rate of growth.

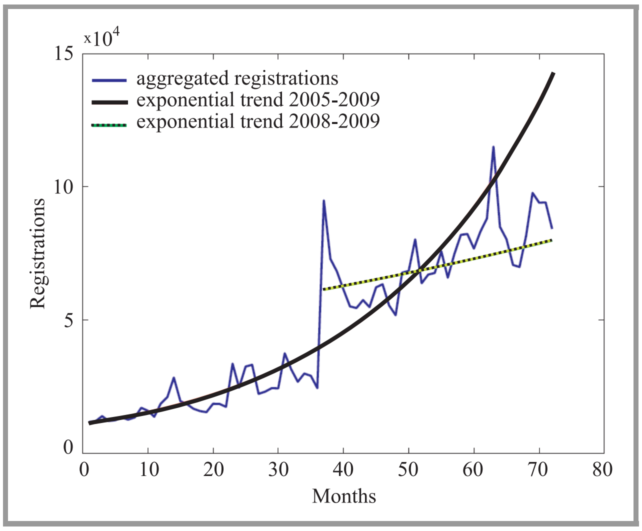

Fig. 7. Exponential trends fitted to deseasonalized registrations: longer (thick) line is trend fitted to the whole 2005–2009 period, shorter (dotted) – 2008–2009 period.

Similarly, the seasonal changes are more regular in last two years (although it can be hardly seen in Fig. 5), so they can be also better identified using shorter period.

The model was constructed by averaging registrations in subsequent months. This way we constructed average registrations sums for January, February, etc., which in connection with the trend provides important information about registration process, and when extrapolated can be used as a simplest prediction (see Fig. 8). Similarly to what can be

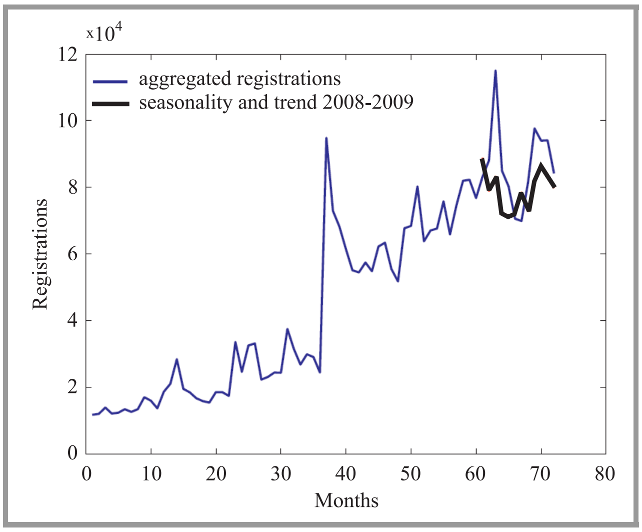

Fig. 8. Registrations forecasted using seasonality and exponential trend for 2008–2009 period.

observed in the Fig. 7, a prediction using the trend and the seasonality fitted to shorter period is much better, in fact it follows the general shape of the line. The greatest discrepancy – in the begging of the predicted period is caused by possible anomaly, which was left unfiltered due to the lack of information – cf. discussion in Section 4 and Fig. 2.

5.2. Residuals Analysis

In order to analyze results of fitting a trend and a seasonality, residuals were analyzed. The graphs in Fig. 9 present quality of fit to learning data and values of residuals. Although the model output follows the general shape of registration process the values of residuals remain significant and, as can be seen in the lower graph in Fig. 9(a), some correlation between values of the modeled process and residuals may be found.

It must be noted that correlation (if it exists) is relatively weak – grouping of points in the lower left side of the plot is not very clear. The presence of correlation suggests that autoregression could be applied to improve the model. To assess the structure of the model an autocorrelation and a partial correlation functions were computed for a process – see Fig. 10. Both ACF and PACF plots decay relatively fast with only first coefficient being significant. Such a shape suggests correlation with the process lagged one interval (month) back, and application of AR(1) model. Values of coefficients for further (10, 11 and 12) intervals remains

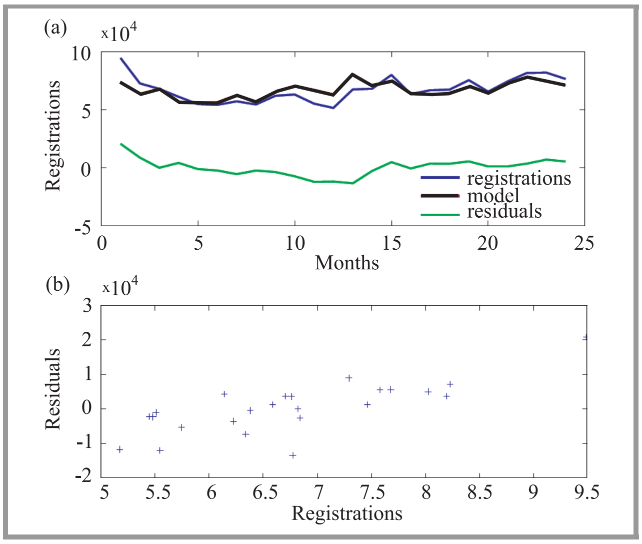

Fig. 9. Quality of fit and residuals for the model using seasonality and trend fitted to 2008–2009 period: (a) quality of fit; (b) values of residuals are plotted against process values.

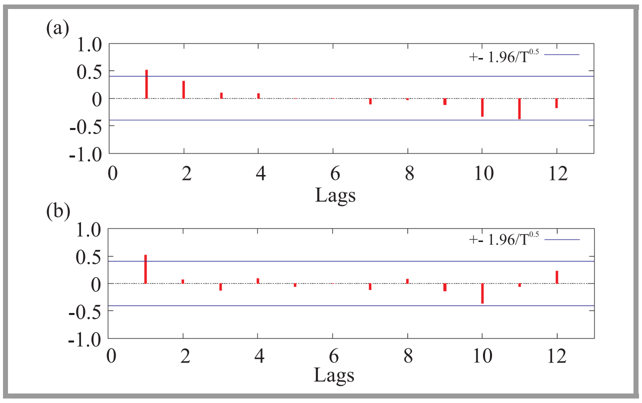

Fig. 10. Autocorrelation – ACF (a) and partial autocorrelation – PACF (b) for residuals of the model using seasonality and trend fitted to 2008–2009 period.

close to significant, which may be caused by some, even weaker correlation, however intervals of 10 or 11 months seem not to be justified by any known property of the process.

5.3. Regressive Modeling

Following analysis in Subsection 5.2, we decided to try to improve the model by applying autoregression to residuals. We started with first order model to begin with the simplest formula and eventually augment it with higher lags after assessing the results. As the model was fitted to the data with trend removed, we neglected intercept and identified only one coefficient. Shorter (2008–2009) data set was used for identification of seasonality and trend, and for computing residuals according to analysis in Subsection 5.1. The resulting AR(1) model proved to be significant, predicted values are shown in Fig. 11. The improvement attained is marginal and visible only in the beginning of the predicted process, however this is implied by the nature of AR(1) model and small values of ACF and PACF coefficients.

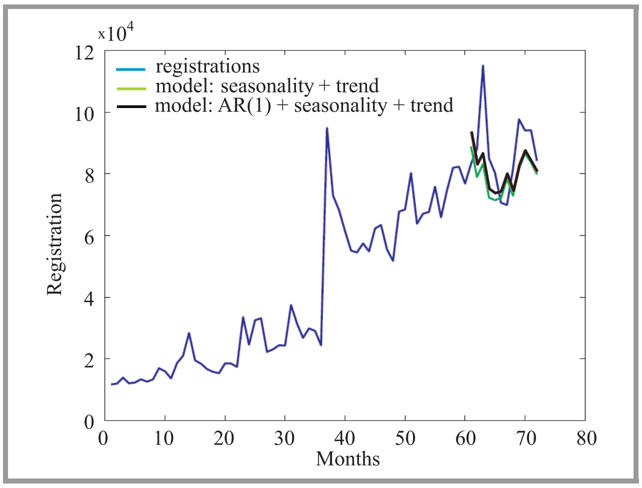

Fig. 11. Prediction by the model augmented with AR(1) versus pure seasonality with trend and data.

To assess the resolving value of the model ACF and PACF of its residuals were computed (see Fig. 12). The analysis of residuals show similarly to earlier results (see Fig. 10), relatively high value of ACF and PACF coefficients for 12th interval, however coefficients for shorter intervals are smaller than in the case of seasonality and trend modeling. Concluding: autocorrelations show that AR(1) model improves model fit with respect to shorter lags, however modeling longer dependencies may be beneficial, especial as the 12th interval has some interpretation in the nature of the analyzed process (yearly correlations caused by yearly rate of payments).

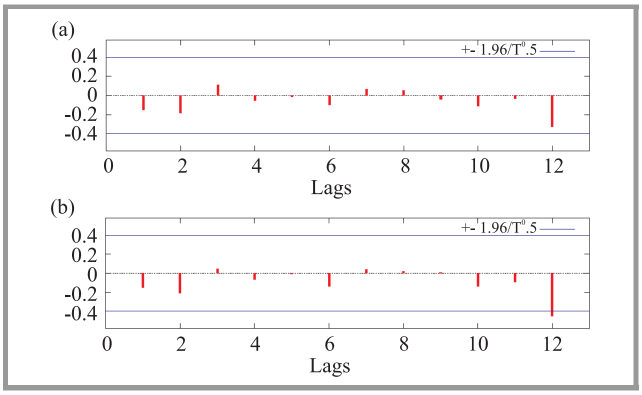

Fig. 12. Autocorrelation – ACF (a) and partial autocorrelation – PACF (b) for residuals of the model using seasonality and trend fitted to 2008–2009 period augmented with AR(1).

To check this hypothesis, AR(12) model consisting of three coefficients: for lags 1, 12 and intercept was fitted. The remaining lags (2-11) were skipped to avoid solving a poorly conditioned problem. The resulting model has significant coefficients, however not to a degree like in the AR(1) case. To assess the fit to the learning data set, Akaike information criterion (AIC) was computed. The application of AIC is

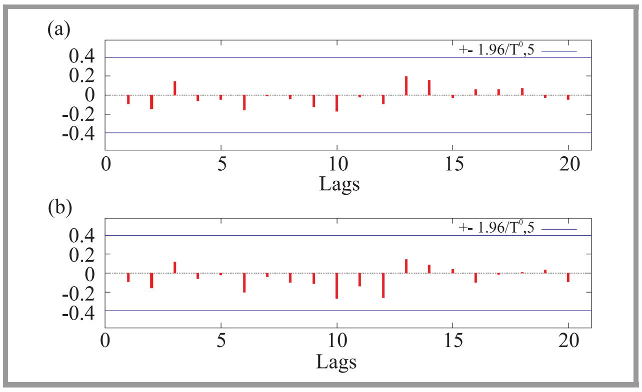

Fig. 13. Autocorrelation – ACF (a) and partial autocorrelation – PACF (b) for residuals of the model using seasonality and trend fitted to 2008–2009 period augmented with AR(12).

reasonable here, as it not only provides measure of fit to the learning data, but also provides correction for complexity of the model. For AR(12) it is a bit better than in case of the AR(1) model (476.5 vs. 481.7), also ACF and PACF (see Fig. 13) show some reduction of coefficients for higher lags. These findings may be contradicted by assessing the quality of prediction – the mean square error for AR(12) model is visibly higher (4165.9 vs. 3361.5). So although the model seems to better reflect the character of learning data its ability of prediction is lower.

To check the possibility of finding better model, we identified and verified a number of models – we tried to test how introduction of longer lags may influence quality of fit and prediction, we also tested effects of using longer period to calculate seasonality (i.e., using again 2005–2009 instead of 2008–2009). To summarize the results we computed two indexes: AIC, and mean square error of prediction to assess the possibility of practical use. The results are presented

Table 1
Comparison of long range models

Model variant	AR lags	Intercept	AIC	Prediction error
Trend period: 2008–2009, seasonality period: 2008–2009				
Trend+seasonality	–	–	–	3689.9
AR(1)	1	–	481.7	3361.5
AR(12)	1, 12	–	476.5	4165.9
AR(12) – 2nd variant	12	–	485.5	4396.4
AR(12) – 3rd variant	1, 12	+	476.9	3797.4
Trend period: 2008–2009, seasonality period: 2005–2009				
Trend+seasonality	–	–	–	3928.1
AR(1)	1	–	496.2	3584.8
AR(11)	1, 11	–	494.4	4140.3
AR(11) – 2nd variant	1, 11	+	496.2	3804.3
AR(10)	1, 10	–	493	4055.6
AR(10) – 2nd variant	1, 10	+	494	4016.8
Trend period: 2005–2009, seasonality period: 2008–2009				
Trend+seasonality	–	–	–	10852.0
Trend period: 2005–2009, seasonality period: 2005–2009				
Trend+seasonality	–	–	–	11098.0

in Table 1. They show that although it is possible to attain better fit to learning data by application of higher order AR model, it does not improve the quality of prediction. Also, as suggested by preliminary analysis using longer period to identify seasonality is ineffective – seasonal changes tend to evolve similarly to trends, however two years period allows to build relatively effective model.

6. Short Range Modeling

Although long range model presented in Section 5 is usually sufficient for making strategic decisions, there are situations when more precise, shorter range predictions are necessary. An example may be assessing resources needed for proper operation of registration databases or planning the advertisement campaign – sometimes even a date of publishing advertisements or billboards may be important. To achieve this goal a completely new model with resolution of days must be built, thankfully the prediction horizon may be reduced, 4 weeks being usually enough. The advantage of a short horizon is that much more data is available. In consequence, models can be better verified. We prepared 18 learning data sets of length 12 weeks selected from period 2008–2010, each of them accompanied by 4 subsequent weeks used for validation. Later, to check properties of models, we shortened learning sets to 4 weeks with validation sets unchanged. Such a construction of data sets allowed to tune 18 models independently and compute mean errors for comparison.

6.1. Model Construction

The model was constructed following the pattern used for long range model (see Subsection 5.1). The most important is seasonality, computed as average number of registrations in subsequent days of a week. Figure 14 shows weekly pattern generated this way, compared with original values of the process. The regularity of the data results in relatively good fit even for such a simple model. The explanation of weekly changes is easier when noted that lower sales

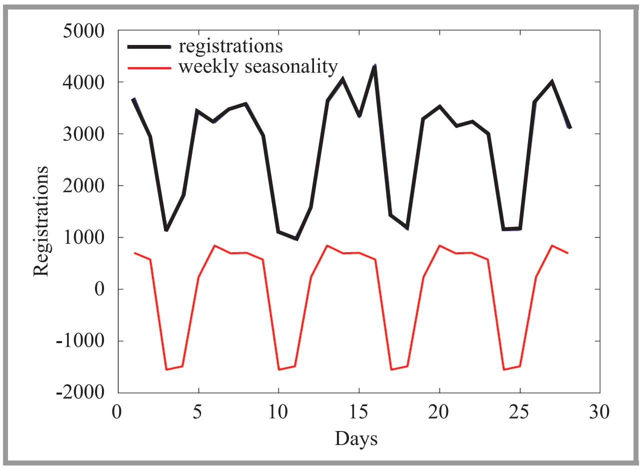

Fig. 14. Weekly seasonality versus registration process.

occurs in weekends. The reason for this may be twofold: first, weeks are scheduled for work – people usually tend to rest during weekends, second (and in fact resulting from the first), bank transfers can not be done on weekends. Payments are only possible by means of other services like, e.g., PayPal or a credit card.

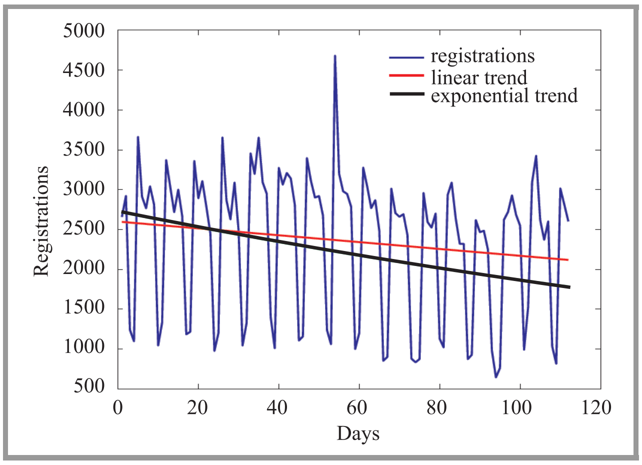

Fig. 15. Fitting trend: learning period of 12 weeks, last 4 weeks is a prediction.

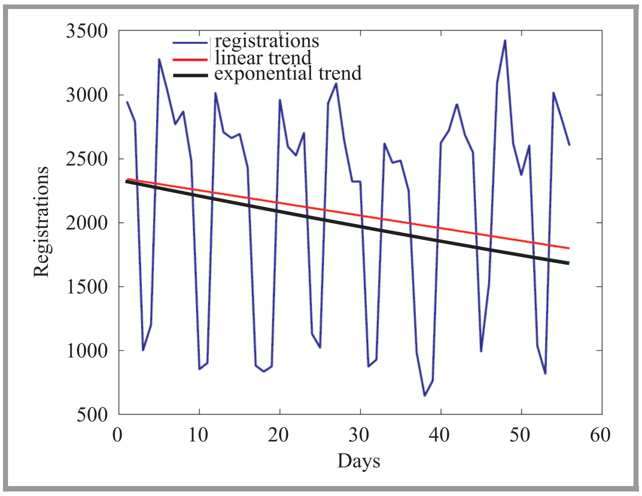

Fig. 16. Fitting trend: learning period of 4 weeks, last 4 weeks is a prediction.

Fitting a trend in a short time horizon is slightly different task than in a timespan of several years. Changes are not so pronounced. For this reason, we tried to use not only previously selected exponential trend, but also a linear one. Another question is a selection of appropriate learning period – there is a danger of unnecessarily introducing long range fluctuations, which are beyond resolution of a short range model. We tried to fit both trends to initially selected learning period (12 weeks) and shortened data set (4 weeks). The results are presented in Figs. 15 and 16 respectively. Observation of graphs allows to find out that longer learning period results in better, a bit damped, estimation. Also, the linear trend performs better, giving more stable prediction.

6.2. Model Validation

Combining seasonality and trend into a single model results in predictions presented in Fig. 17 for learning period of 12 weeks and 4 weeks (Fig. 18). Parts (a) figures show prediction compared to observed reservations while (b) two ACF and PACF plots respectively, in both cases the prediction is calculated for 25-03-2008 to 21-04-2008 being typical period for all of 18 analyzed samples.

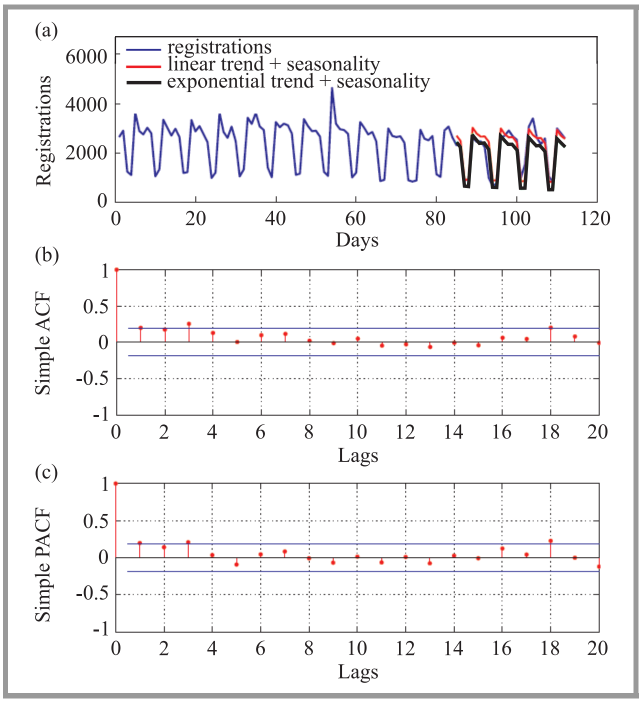

Fig. 17. Prediction for 4 weeks using seasonality and trend, learning period of 12 weeks: (a) prediction itself; (b) ACF, and (c) PACF of residuals for linear trend and seasonality.

Results are surprisingly good, especially in case of 12 week learning period and linear trend. Of course, it is impossible to predict some rapid, individual changes like e.g. in the second part of prediction, however, the fact, that all coefficients in the residuals ACF (see Fig. 17(b)) are reduced, proves the quality of proposed model. Such a shape of autocorrelation suggests that application of autoregressive models to improve prediction would be nearly impossible – and it was indeed the result of our trials. On the other hand, the PACF graph of the model tuned to shorter period of data (see Fig. 18(c)) shows some interesting properties – although coefficients for most of lags are highly reduced, the lag 14 coefficient is significant, suggesting some dependence on the span of two weeks. This hypothesis seems to be understandable – the presence of such a cycle may be somehow explained (e.g., investors may observe market in one week and then take decisions). However, building 14th order autoregressive model to encompass this is hardly feasible (and it proved to be), especially when confronted with results of modeling using 12 weeks of learning data, when this problem is overcome by averaging over longer period.

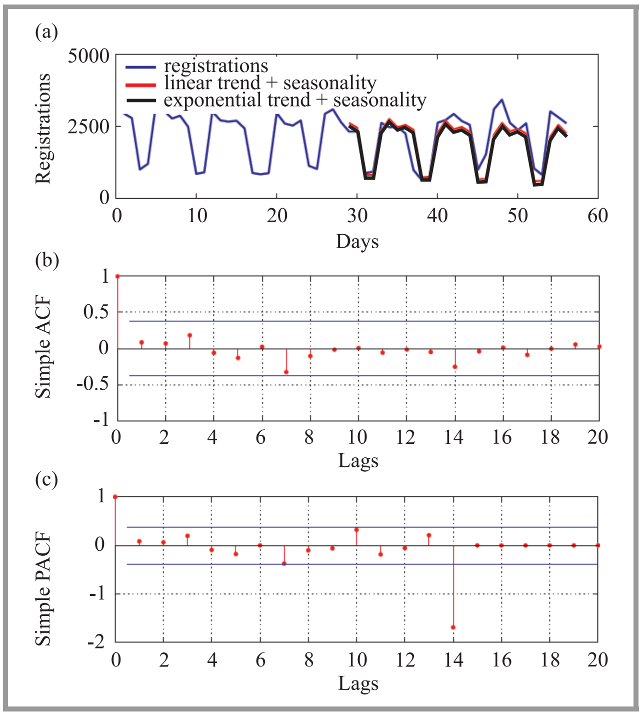

Fig. 18. Prediction for 4 weeks using seasonality and trend, learning period of 4 weeks: (a) prediction itself; (b) ACF, and (c) PACF of residuals for linear trend and seasonality.

Table 2
Comparison of short range models

Learning period	Trend	Error of 18 predictions
1 month	linear	166.12
1 month	exponential	178.45
3 months	linear	134.59
3 months	exponential	139.15

To summarize: as results for the model constructed of seasonality and linear trend tuned to longer period of data was sufficient to describe most of short range properties of registration process, and attain precision of approx. 15%, we refrained from further refinement. The results in the form of mean square error of 18 cases for all analyzed variants are presented in Table 2.

7. Composite Modeling

Encouraged by promising results acquired with long and short range models, we decided to try to construct a model, which while having long range (possibly one year) capability will allow prediction with high resolution – possibly of one day like the short range model. Such a model can be useful for making some decisions based on precise forecast of registrations, it can also provide some important information on the nature of the analyzed process. The possibility of building such a model is mostly grounded by the fact of relatively high regularity of weekly cycles what was shown in Section 6.

7.1. Model Construction

The core of the model is monthly registration sums computed by means of the long range model. The best version of the model i.e. with calculation of seasonality and trend using two years data and AR(1) model was used. Monthly sums are interpolated linearly over subsequent days of a month, as it was shown that the linear trend performs better in the short range model. Obtained this way, monthly trend is then modified with weekly seasonality calculated in similar way, as for the short range model but independently for subsequent months. This way, different shape of weekly cycle (mostly amplitude) is taken into account. During initial evaluation we found out that the amplitude of weekly cycles changes in subsequent years – typically it grows, when number of registrations grows. This phenomenon can not be modeled by summation of a trend and seasonality – to encompass it we introduced a multiplicative factor – amplitude growth rate.

7.2. Model Validation

The same, as in the previous experiments learning data consisted of daily registrations in years 2008–2009, while data from 2010 was used for validation. Five variants of the model were compared, they differed in the way of calculation of the following components:

– weekly seasonality: for the whole period or one year selected,

– amplitude growth rate: none, monthly or annual.

The reason for shortening data period used for weekly seasonality computation was the occurrence of the above mentioned changes in the amplitude of cycles. Two variants of amplitude growth rate were calculated to identify its nature: eventually it can be stated that the growth of amplitude may be seen as long range process correlated with general (yearly) trend.
Validation showed that all five models behave surprisingly well, describing most of significant properties of the data. The most important is the ability to follow general trend and

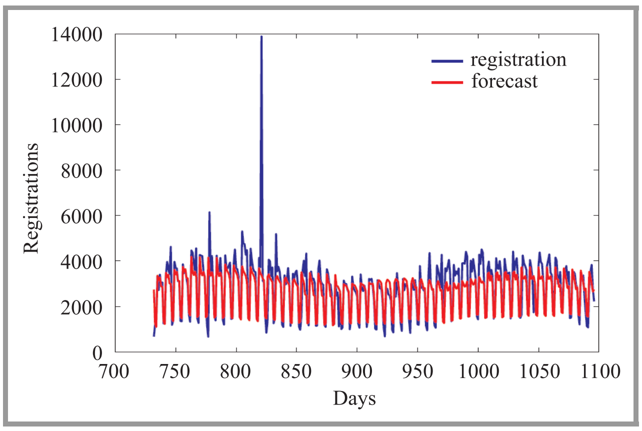

Fig. 19. Prediction for year 2010 using composite model with annual amplitude growth rate.

to model seasonal variations of weekly amplitude. Modeling of the last property is to some extent improved by introducing multiplicative component – annual rate of growth – in the most successful model (see Fig. 19).
The performance of all models is summarized in Table 3. Although the results are very good, it must be noted that

Table 3
Comparison of composite models

Weekly seasonality	Growth rate	Mean square error
2008–2009	none	820.95
2008	none	841.11
2009	none	831.72
2008–2009	monthly	836.42
2008–2009	yearly	816.39

still some periods when customers behave differently than usual (e.g., rise in the beginning of autumn 2010), and anomalies cannot be predicted. To analyze performance better ACF and PACF of residuals were computed (see Fig. 20). The results are difficult to interpret and probably need the further analyses. What can be stated now is that not all coefficients of ACF in the range of 1 to 50 days are sufficiently reduced, which may suggest presence of some unmodeled dependencies. Also, the PACF graph does not decay smoothly – there are some lags of length between 180–240 days, which have significant coefficients. The 6 month (approx. 180 days) lag may be to some extent attributed to two periods of higher sales observed in every year while longer may result from irregularities caused by external factors.

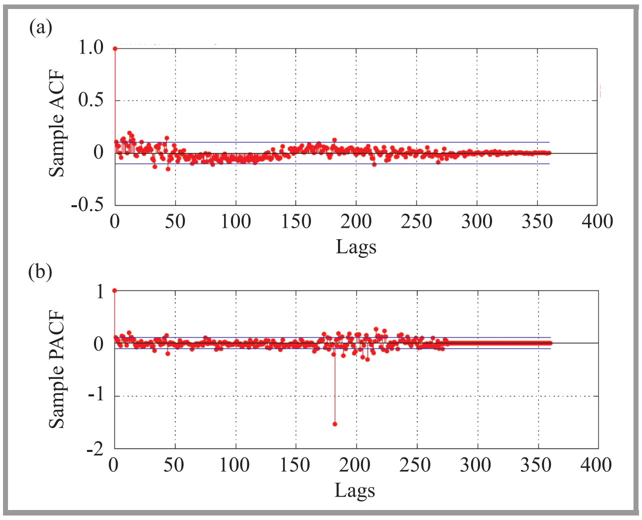

Fig. 20. Autocorrelation (a) and partial autocorrelation (b) for residuals of the composite model with annual amplitude growth rate.

Another question implied by this analysis is the presence of long range dependence in the registration process. The autocorrelations (also these computed for long range model) can not answer this question clearly – first of all, the number of samples is relatively small. Nevertheless, we tried to estimate Hurst coefficient for the process by fitting fractional Brown motion process. We ended with Hurst coefficient of 0.8 and relatively poor fit. Our supposition is that the long range dependence in the registration process is possible, however, it is likely that it is implied by other socio-economic variables, e.g., economic cycles to name most obvious one, which in turn are known to be long range dependent.

8. Conclusions

We have analyzed data and proposed models for various time scales. The most important outcome of these analyses is in our opinion identification of periodic nature of registration process. The periodicism has two scales – shorter, connected with weekly cycle and longer, visible as two periods of lower sales during the year. Another important part of the process is a trend, which in long range may be best modeled by exponential curve. These components were used to build models proposed, which proved to be precise enough for planning marketing strategies or sizing hardware.
There are also factors we do not cover in our models – mostly connected with external variables, which influence registrations. We roughly identified two such variables: one is general socio-economic situation and the second are prices. Both of them are difficult to comprehend, especially in the case of prices it is difficult to observe strategies of all dealers selling domains. However, we plan to analyze the influence of external factors deeper and to input them into the model, possibly in an aggregated form using indexes and statistics. We hope to solve the problem of nonstationarity this way and eventual long range dependence, which we could observe in the data.

References

[1] M. Mueller, "Toward an economics of the domain name system", in *Handbook of Telecommunications Economics: Technology Evolution and the Internet*, M. E. Cave, S. K. Majumdar, and I. Vogelsang Eds., vol. 2, North-Holland, 2005, pp. 443–487.

[2] M. Jindra, "The market for Internet domain names", in *Proc. 16th ITS Regional Conf.*, Porto, Portugal, 2005 [Online]. Available: http://userpage.fu-berlin.de/∼jmueller/its/conf/porto05/papers/Jindra.pdf, accessed 12.08.2011.

[3] "The Secondary Market For Domain Names", *OECD Report DSTI/ICCP/TISP(2005)9/FINAL*, 2005.

[4] M. Zook, "Study of the Factors Behind the Demand for Country Code Domain Names", ZookNIC Internet Intelligence [Online]. Available: http://www.aptld.org/pdf/APTLD-Bangkok-07-Report-FINAL.pdf, 2007, accessed: 12.08.2011.

[5] P. Salvador and A. Nogueira, "Analysis of the Internet domain names re-registration market", *Procedia Comp. Science*, p. 325–335, 2011.

[6] K. Lasota and K. Kozakiewicz, "Analysis of the similarities in malicious DNS domain names", *Commun. Computer Inform. Sci.*, vol. 187, pp. 1–6, Springer, 2011.

[7] J. Wu, X. Li, X. Wang, and B. Yan, "DNS usage mining and its two applications", in *Proc. Sixth Int. Conf. Digit. Inform. Manag. ICDIM 2011*, Melbourne, Australia, 2011, pp. 54–60.

[8] T. Lindenthal, "Valuable Words: Pricing Internet Domains", Social Science Research Network, July 14, 2011 [Online]. Available: http://ssrn.com/abstract=1885465, accessed 09.01.2012.

[9] A. Tajirian, "Statistical Models for Market Approach to Domain Name Valuation", 2010 [Online]. Available: http://www.domainmart.com/news/Statistical_Models_for_Market_Approach_to_Domain_Name_Value.pdf, accessed 09.01.2012.

[10] H. Simon, *Price Management*. North-Holland, 1989.

[11] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. Holden-Day, 1970.

[12] S. Makridakis, S. C. Wheelwrighth, and R. J. Hyndman, *Forecasting: Methods and Applications*. Willey, 1998.

[13] J. Haywood and J. Randal, "Modelling seasonality and multiple structural breaks. Did 9/11 affect visitor arrivals to NZ?", *Research report 08/10*, University of Wellington, New Zealand 2008 [Online]. Available: http://msor.victoria.ac.nz/twiki/pub/Main/ResearchReportSeries/mscs08-10.pdf, accessed 12.08.2011.

[14] M. K. Francke, "Repeat Sales Index for Real Estate Prices: a Structural Time Series Approach", *Technical Working paper 2009-0*, Ortec Finance Research Center, 2009.

[15] D. A. Pierce, "A survey of recent developments in seasonal adjustment", *The American Statistician*, vol. 34, no. 3, 1980.

[16] Ch. I. Plosser, "A time series analysis of seasonality in econometric models", in *Seasonal Analysis of Economic Time Series*, A. Zellner, Ed. NBER, 1979, pp. 365–410.

[17] B. Sheharyar and D. Geltner. "Estimating real estate price movements for high frequency tradable indexes in a scarce data environment", *The Journal of Real Estate Finance and Economics*, 2010 [Online]. Available: http://www.springerlink.com/content/uxl087168u781086/, accesed: 09.08.2011.

[18] L. Ferrara and D. Guagan, "Fractional seasonality: Models and Application to Economic Activity in the Euro Area", *Working paper*, Eurostat 2006 [Online]. Available: http://epp.eurostat.ec.europa.eu, accessed 12.08.2011.

[19] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "STL: a seasonal trend decomposition procedure based on loes", *J. Official Statistics*, vol. 6, no. 1, pp. 3–73, 1990.

[20] M. Yogo, "Measuring Business Cycles: Wavelet analysis of economic time series", *Economics Letters*, vol. 100, no. 2, 2008, [Online]. Available: http://ssrn.com/abstract=1120482, accessed 10.08.2011.

Piotr Arabas received his Ph.D. in Computer Science from the Warsaw University of Technology, Poland, in 2004. Currently, he is assistant professor at Institute of Control and Computation Engineering at the Warsaw University of Technology. Since 2002 with Research and Academic Computer Network (NASK). His research area focuses on modeling computer networks, predictive control and hierarchical systems.

E-mail: parabas@ia.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

E-mail: Piotr.Arabas@nask.pl
Research Academic Computer Network (NASK)
Wąwozowa st 18
02-796 Warsaw, Poland

Przemysław Jaskóła received his M.Sc. in Computer Science from the Warsaw University of Technology, Poland, in 1999. Currently, he is a Ph.D. student in the Institute of Control and Computation Engineering at the Warsaw University of Technology. Since 2005 with Research and Academic Computer Network (NASK). His research area focuses on hierarchical optimization and computer networks.

E-mail: Przemyslaw.Jaskola@nask.pl
Research Academic Computer Network (NASK)
Wąwozowa st 18
02-796 Warsaw, Poland

Mariusz Kamola received his Ph.D. in Computer Science from the Warsaw University of Technology, Poland, in 2004. Currently, he is assistant professor at Institute of Control and Computation Engineering at the Warsaw University of Technology. Since 2002 with Research and Academic Computer Network (NASK). His research area focuses on economics of computer networks and large scale systems.

E-mail: mkamola@ia.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska 15/19
00-665 Warsaw, Poland

E-mail: Mariusz.Kamola@nask.pl
Research Academic Computer Network (NASK)
Wąwozowa st 18
02-796 Warsaw, Poland

Michał Karpowicz received his Ph.D. in Computer Science from the Warsaw University of Technology (WUT), Poland, in 2010. Currently, he is an Assistant Professor at Research and Academic Computer Network (NASK). His research interests focus on game theory, network control and optimization.

E-mail: Michal.Karpowicz@nask.pl
Research Academic Computer Network (NASK)
Wąwozowa st 18
02-796 Warsaw, Poland