Best Method Among three Trend Analysis Methods and Its Forecasting

Abstract

We study the trend analysis methods, namely (1) least square method, (2) trend quadratics method, and (3) exponential trend method, to get which one is the best in forecast the human development index (HDI) time series data. Here, we used mean absolute percentage error (MAPE) in comparing the three methods to obtain the eligible one. A simulation is given then using on the HDI data from Cilacap regency. The result showed that the smallest MAPE (on the trend quadratics method) is 0.19%. Due to this value of the MAPE, we then choose the trend quadratics method as the best model. Furthermore, we used this method to forecast the simulation HDI data, and the result showed that the values of the forecasting are 70.6, 70.7, and 70.7 for period 2022 to 2024, respectively. These result indicate that the trend of the HDI in Cilacap tend to be slowly increase.

Keywords: HDI, forecasting, trend analysis, quadratic trend

1. Introduction

In the theory of statistics, we note that there are three important concepts of the hypothesis testing in rejecting null hypothesis, namely (1) probability error type I (α), (2) a probability error type II (β) and (3) a power {Wackerly et al.[1]). One of the useful of thing is power, it is an eligible method to test a parameter distribution on one-side or two-side hypothesis. Here, we note that the statistics analysis is not only hypothesis testing but also exploratory data analysis, time series analysis, survival analysis, etc. Therefore, we then interested to study trend analysis, namely (1) least square method, (2) trend quadratics method, and (3) exponential trend method to forecast the human development iindex (HDI) time series data in Cilacap regency, period 2010 to 2021. So, we assumed that this is as one of the a real application (forecasting) on time series method. To understanding more, we firstly review about the power of the test method in testing parameter. We note here that many authors already studied the power in testing parameter shape on some distributions and or hypothesis testing, such as, Pratikno [2] and Khan and Pratikno [3], studied the power in testing intercept with non-sample prior information (NSPI). They used the probability integral of the cumulative distribution function (cdf) of the distributions to calculate the power. Moreover, Khan et al. [4] used the power to compute the cdf of the bivariate noncentral *F* (BNCF) distribution in multivariate and multiple regression models. In this research context, we also noted that many authors, such as Khan [5], Khan and Saleh [6,7,8], Khan and Hoque [9], Saleh [10], Yunus [11], and Yunus and Khan [12,13,14], have contributed to the research of the power in the estimator and hypothesis area.

In the context of the real application, we used secondary data from statistics bureau of Indonesia (BPS), that is HDI, from Cilacap regency, period 2010 to 2021. This data tend to

yearly increase. We can see that the HDI in 2020 is 69.95 and HDI in 2021 is 70.42 (increase 0.47). But, in the context of the time series data, there are two types of methods, namely the qualitative and quantitative forecasting methods. One of the quantitative forecasting method is trend analysis (see, Maryati [15]). Following Makridakis, et al. [16], there are three type on the trend analysis method, namely (1) least square method $\left(Y = a + bX\right)$, (2) trend quadratics method $\left(Y = a + bX + cX^2\right)$, and (3) exponential trend method $\left(Y = ab^X\right)$. Following Yulfida, [17], these methods are well-known and an accurate. Here, we present some previous research in the context of forecasting and its simulation such as Suharyadi and Purwanto, [18], Saputro and Purwanggono [19], Fahmi and Muslim [20], Siagian [21], Pahala [22], Purba [23], Rachman [24], Billah, et. al. [25], Farida, et. al [26], and Robial [27], Sagar and Najam [28], and Setiawan and Hakim [29].

From the background above and the pattern of the plot of the HDI data (increasing trend), we then focused to study HDI as time series data using trend analysis methods. The paper deals with HDI calculation (simulation data) to identify which one be the best model among three type of the trend analysis methods, namely least square method, trend quadratics method and exponential trend method.

In this paper, we then presented the introduction in Section 1. The research methodology is given in Section 2. The result are obtained in Section 3, and the conclusion is provided in Section 4.

2. Methodology

There are three steps in comparing among trend analysis methods (least square method, trend quadratics method, and exponential trend method) as follow (1) overviewed the literature theory about MAPE, (2) computed the real actual error between the forecasting data and actual data, and (3) choose the minimum MAPE. Furthermore, we simulate using HDI data of the statistics bureau of Indonesia (BPS) Cilacap regency, from 2010 to 2021, at Table 1.

Table 1. HDI Data of Cilacap 2010-2021

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
HDI	64.2	64.7	65.7	66.8	67.3	67.8	68.6	68.9	69.6	70.0	69.9	70.4

The steps of analysis data are given as follow: (1) identifying trend of the *time series plot* data HDI, (2) forecasting the data using three methods, (1) least square method, (2) trend quadratics method, and (3) exponential trend, (3) ccomputed the MAPE and choose the lowest MAP, (4) using MAPE we then test the best model from the three methods, the least square method, trend quadratics method, and exponential trend, (5) by choosing the best model in (4) , we then forecast the data from 2022 to 2025.

We see from Table 1 that the HDI data tend to be positive trend. It means that the suitable method is *trend analysis*, We therefore used three methods of the trend analysis, namely (1) least square method, (2) trend quadratics method, and (3) exponential trend, in forecasting the data. The best model is then chosen by e choosing the lowest mean absolute percentage error (MAPE), as $MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{X_t - F_t}{X_t}\right| x100\%$, among them. Here, F_t is the value of the forecasting at period t, and X_t is the actual data at period t. Here, (1) if the MAPE < 10%, we said that the model is very good, (2) if the MAPE between 10%-20%, it means that the model is good, and (3) if the MAPE lie between 20%-50%, we suspect that the model is enough, otherwise is not good.

3. Result

3.1. Least Square Method

In this Section, we estimated the coefficients *a* and *b* of the least square method, $Y = a + bX$, using $a = \frac{\sum_{i=1}^{n=12} Y_i}{n}$ and $b = \frac{\sum_{i=1}^{n=12} X_i Y_i}{\sum_{i=1}^{n=12} X_i^2}$. In this concept, we have to create the code number for *X* from 2010 to 2021, that are -11, -9, -7, -5, -3, -1, 1, 3, 5, 7, 9, and 11, for computing the coefficients *a* and *b* as follow, respectively, $a = 813.9/12 = 67.9$ and $b = 165.02/572 = 0.29$. Here, the value of the MAPE is then given as

$$\text{MAPE} = \frac{\sum_{i=1}^{n=12} \left| \frac{X_i - F_i)}{X_i} \right|}{n} x100\% = \frac{0.062}{12} x100\% = 0,5\%.$$

3.2. Trend Quadratics Method

Based on the Section 3.1., we then defined the trend quadratics method. In this concept, the model is written as $Y = a + bX + cX^2$. Here, the estimate model is computed as

$$a = \frac{\sum_{i=1}^{n=12} Y_i - c\sum_{i=1}^{12} X_i^2}{n} = \frac{813.9 - (-0.008)(572)}{12} = 68.2, \qquad b = \frac{\sum_{i=1}^{n=12} X_i Y_i}{\sum_{i=1}^{n=12} X_i^2} = \frac{165.02}{572} = 0.3$$

and

$$c = \frac{\sum_{i=1}^{12} X_i^2 \sum_{i=1}^{n=12} Y_i - n\sum_{i=1}^{12} Y_i X_i^2}{\left(\sum_{i=1}^{12} X_i^2\right)^2 - n\sum_{i=1}^{12} X_i^4} = \frac{(572)(813.9) - 12(38615.7)}{(572)^2 - 12(48620)} = -0.008.$$

Hence, the model of the quadratics method is obtained as $Y = 68.2 + 0.3X - 0.008X^2$. If X increases 1, the value of the Y increases 0.282. Furthermore, the value of MAPE is given as

$$\text{MAPE} = \frac{\sum_{i=1}^{n=12} \left| \frac{X_i - F_i)}{X_i} \right|}{n} x100\% = \frac{0.023}{12} x100\% = 0.19\%.$$

3.3. Exponential Trend Method

We presented here, the model of the exponential trend method, that is $Y = ab^X$. To estimate this model, we estimated the coefficient *a* and *b* using the formula $\log(a) = \frac{\sum_{i=1}^{n=12} \log(Y_i)}{n}$

and $\log(b)=\dfrac{\sum_{i=1}^{n=12} X_i \log(Y_i)}{\left(\sum_{i=1}^{n} X_i^2\right)}$. It is easy to compute the estimated coefficients a and b as follow, respectively $\log(a)=\dfrac{\sum_{i=1}^{n=12}\log(Y_i)}{n}=\dfrac{21.97}{12}=1.8\Rightarrow a=67.8$ $\log(b)=\dfrac{\sum_{i=1}^{n=12} X_i \log(Y_i)}{\left(\sum_{i=1}^{n} X_i^2\right)}=\dfrac{1.06}{572}=0.002\Rightarrow b=1.0.$

Finally, we then get the model of the exponential trend method as $Y=(67.8)(1.004)^X$. It means that if the data year X is zero, the forecasting of Y , will be 67.8, otherwise if year X increase 1, the value of the Y increases $(67.8)(1.0)^1$. In this context of the exponential trend method, the MAPE is obtained as

$$\text{MAPE}=\frac{\sum_{i=1}^{n=12}\left|\dfrac{X_i-F_i)}{X_i}\right|}{n}x100\%=\frac{0.064}{12}x100\%=0,53\%$$

From the above MAPE, the lowest MAPE is 0.19%. Due to this situation, we then conclude that the best model is the trend quadratics method. Furthermore, we used this method to forecast the HDI data of the Cilacap regency for period 2022-2024 at Table 2.

Table 2. The result of the HDI forecasting

Year	X	Forecasting	Actual	Error
2022	13	70.6	70.99	0.39%
2023	15	70.7		
2024	17	70.7		

We see from Table 2 that the trend of the HDI data in Cilacap tend to be slowly increase.

4. Conclusion

In this research, we compare among the three methods on the trend analysis, namely (1) least square method, (2) trend quadratics method, and (3) exponential trend method, obtain the best model among them. Here, a simulation is conducted using the data of the HDI (from Cilacap regency, period 2010 to 2021). This is as a real application in planning and others analysis related to HDI in Cilacap. Using the three methods above, we compute the estimated model and MAPE to get the best model. The result showed that the smallest MAPE (on the trend quadratics method) is 0.19% . Due to this value of the MAPE, we then choose the trend quadratics method as the best model. Furthermore, we forecast the data using the trend quadratics method, that are 70.6, 70.7, and 70.7, for period 2022 to 2024, respectively. These result indicate that the trend of the HDI in Cilacap always slowly increase.

Acknowledgment

I thankfully to the LPPM UNSOED for providing me granting of research

References

[1] Wackerly, D.D., Mendenhall III, W, R. L.Scheaffer. Mathematical Statistics With Application, 7th Ed. Thomson Learning, Inc., Belmont, CA, USA, 2008, https://books.google.co.id/books/about/Mathematical _Statistics_with_Application.html?id=lTgGAAAAQBAJ&redir_esc=y

[2] Pratikno, B. Test of Hypothesis for Linear Models with Non-Sample Prior Information. Unpublished PhD Thesis, University of Southern Queensland, Australia, 2012,https://research.usq.edu.au/download/fd27eeffb76c7e0a66d0e8e885a132043def7de51f760e23e7fac7d511f4348a/904968/Pratikno_2012_whole.pdf.
[3] Khan, S., Pratikno, B. Testing Base Load with Non-Sample Prior Information on Process Load. Statistical Papers, 54 (3), 605-617, 2013, DOI: 10.1007/s00362-012-0451-40.81.
[4] Khan, S., Pratikno, B., Ibrahim, A.I.N, Yunus, R.M. The correlated bivariate noncentral *F* distribution and Its application. Communications in Statistics—Simulation and Computation, 45 3491–3507, 2016, DOI: 10.1080/03610918.2014.999084.
[5] Khan, S. Estimation of the Parameters of two Parallel Regression Lines Under Uncertain Prior Information. Biometrical Journal, **44**, 73-90, 2003, doi.org/ 10.1002/bimj.200290017
[6] Khan, S. Shrinkage estimators of intercept parameters of two simple regression models with suspected equal slopes. Communications in Statistics - Theory and Methods, **37**, 247-260, 2008, doi.org/10.1080/ 03610920701648961
[7] Khan, S., Saleh, A. K. Md. E. Shrinkage pre-test estimator of the intercept parameter for a regression model with multivariate Student-t errors. Biometrical Journal,**39**,117,1997,doi.org/10.1001/bimj.4710380202
[8] Khan, S., Saleh, A. K. Md. E. On the comparison of the pre-test and shrinkage estimators for the univariate normal mean. Statistical Papers, **42**(4), 451-473, 2001, doi.org/10.1007/s003620100073
[9] Khan, S., Hoque, Z. Preliminary test estimators for the multivariate normal mean based on the modified W, LR and LM tests. Journal of Statistical Research, Vol 37, 43-55, 2003, http://www.isrt.ac.bd/node/150
[10] Saleh, A. K. Md. E. Theory of preliminary test and Stein-type estimation with applications. John Wiley and Sons, Inc., New Jersey, 2006, DOI:10.1002/0471773751
[11] Yunus, R.M. Increasing power of M-test through pre-testing. Unpublished PhD Thesis, University of Southern Queensland, Australia, 2010, https://www.cambridge.org/core/journals/bulletin-of-the-australian-mathematical-society/article/increasing-power-of-mtest-through-pretesting/862C81029B3D370918414B10520B7CC1
[12] Yunus, R.M., Khan, S. Test for intercept after pre-testing on slope a robust method. In: 9th Islamic Countries Conference on Statistical Sciences (ICCS-IX): Statistics in the Contemporary World - Theories, Methods and Applications, 2007, https://www.researchgate.net/publication/228779650_Test_for_intercept_after_pre-testing_on_slope-a_robust_method
[13] Yunus, R.M., Khan, S. Increasing power of the test through pre-test a robust method. Communications in Statistics-Theory and Methods, **40**, 581-597, 2011a, https://doi.org/10.1080/03610920903350572
[14] Yunus, R.M., Khan, S. M-tests for multivariate regression model. Journal of Nonparamatric Statistics, **23**, 201-218, 2011b, https://doi.org/10.1080/ 10485252.2010.503896
[15] Maryati. (2010). *Statistika Ekonomi dan Bisnis. Edisi Revisi*. Cetakan Kedua. Yogyakarta: UPP AMP YKPN.
[16] Makridakis, S., Wheelwright, S. C., and McGee, V. E. (2003). Metode dan Aplikasi Peramalan (Jilid 1 ed.). Jakarta: Binarupa Aksara. https://onesearch.id/Record/IOS2862.UNMAL000000000016557
[17] Yulfida, T. (2012). Analisis Trend Pendapatan PT. Angkasa Pura 1 (PERSERO) Bandar Udara International Adi Soemarno Surakarta. (Skripsi, Fakultas Ekonomi, Universitas Sebelas Maret: Surakarta).https://digilib.uns.ac.id/dokumen/detail/27428/Analisis-Trend-Pendapatan-PT-Angkasa-Pura-I-Persero-Bandar-Udara-Internasional-Adi-Soemarmo-Surakarta
[18] Suharyadi, dan Purwanto. (2007). Statistika untuk Ekonomi dan Keuangan Modern Edisi 2. Jakarta: Salemba Empat. https://onesearch.id/Record/ IOS5392.ai:slims-20516/TOC
[19] Saputro, A., dan Purwanggono, B. (2016). Peramalan Produksi Semen Dengan Metode Exponential Smoothing pada PT. Semen Indonesia. *Industrial Engineering Online Journal, 5(4).* 4. https://ejournal3.undip.ac.id/index.php/ieoj/article/view/14052

[20] Fahmi, A. K., dan Muslim, M. A., (2017). Peramalan Nilai Inflasi Kesehatan di Kabupaten Semarang dengan Metode Kuadratik. FMIPA Matematika. Universitas Negeri Semarang. https://www.unaki.ac.id/ejournal/ index.php/komputaki/article/view/149

[21] Siagian, Y. (2018). Implementasi Metode Least Square untuk Peramalan Pertumbuhan Penduduk pada Kabupaten Asahan. *System Informasi. STMIK ROYAL*. 275-380

[22] Pahala, F. (2019). Prediksi Lalu-lintas Penumpang Bandar Udara Soekarno-Hatta dengan Teknik Time Series Trend Forecasting. *Jurnal Penelitian, 4(3),* 1-10.

[23] Purba, A. (2015). Perancangan Aplikasi Peramalan Jumlah Calon Mahasiswa Baru yang Mendaftar Menggunakan Metode Single Exponential Smoothing (Studi Kasus: Fakultas Agama Islam UISU). *Jurnal Riset Komputer (JURIKOM), 2,* 8-12.

[24] Rachman, R. (2018). Penerapan Metode Moving Average dan Exponensial Smoothing pada Peramalan Prouksi Industri Garment. *Jurnal Informatika, 5,* 211-220

[25] Billah, B., King, M. L., Snyder, R. D., and Koehler, A. B. (2006). Exponential smoothing model selection for forecasting. *International journal of forecasting*, *22*(2), 239-247.

[26] Farida, Y., Sulistiani, D. A., dan Ulinnuha, N. (2021). *Peramalan Indeks Pembangunan Manusia (IPM) Kabupaten Bojonegoro Menggunakan Metode Double Exponential Smoothing Brown*. Teorema: Teori dan Riset Matematika.

[27] Robial, S. M. (2018). Perbandingan Model Statistik pada Analisis Metode Peramalan Time Series (Studi Kasus: PT. Telekomunikasi Indonesia, TBK Kandatel Sukabumi). *Jurnal Ilmiah SANTIKA, 8.*

[28] Sagar, A. D., dan Najam, A. (1998). The Human Development Index: A Critical Review. *Ecological Economic, 249-264.*

[29] Setiawan, M. B., dan Hakim, A. (2008). Indeks Pembangunan Manusia Indonesia. *Jurnal Ekonomia, 9(1),* 18-26.