JOURNAL OF ECONOMIC SCIENCES: THEORY AND PRACTICE, V.80, # 2, 2023, pp. 4-13

FORECASTING OF ENGEL CURVE COMPONENTS WITH THE APPLICATION OF ARIMA METHOD

ASIF ALIYEV

Control Systems Institute, Ministry of Science and Education Republic of Azerbaijan, Baku, AZ1141
E-mail: msc.aaliyev@gmail.com

Received: September 15; accepted November 30, 2023; published online December 25, 2023

ABSTRACT
Engel curve coefficient is still actual to obtain some information about wellbeing and poverty level in the country based on structure of households' budget expenses, more exactly through the ratio of food expenses on overall expenses. There exist numerous researches in this course covering household demand behavior effects on Engel curve. In the current paper we consider only middle-income group in population and ignore other factors affecting as income interval changes. The question of interest is the prediction of Engel curve components as an indicator of middle social group well-being. The forecast of Engel curve components for the next few years is carried out using ARIMA model.

Keywords: Engel curve components, middle social group, household, ARIMA model, time series forecasting.

Jel Classification: C32, C53, D63

INTRODUCTION
According to Ernst Engel (1857) the poorer a family, the more portion of total expenditures is to spend on food. Engel curves were widely studied by the application of different econometric methods for different consumption items. Banks et al. (1997) studied Engel curve and consumer demands employing non-parametric analyses. Blundell et al. (1998) studied the consumer demands with the application of semi-parametric methods and outlined the different demographical profiles of households based on Engel curve. You (2003) used robust and standard models in his study where food, transportation, and other expenditures were considered with Engel functions. The role of household behavior effects as consumer on individual preferences is widely reviewed in the work of Solomon (2006), aside the economics. Levie & Xu (2007) found out that how households adjust their consumption on food, housing, education, and other stuff to handle budget expenditures.

Kumar & et.al (2008) showed that food consumption deprivation index has very little correlation with the traditional measures of poverty. They raised a question how to combine Engel curves for some other vital goods such as health services, education, water, energy, etc. into a single index of poverty. Almas (2012) estimated households Engel curves for nine different countries, and observed stability of the Engel law across the countries. James & et.al (2012) estimated food elasticities at different levels of expenditure indicating that as income and overall food expenditure increases, expenditure for foods will increase at a declining rate in line with Engel's law. Çağlayan (2012) estimated Engel curves for food and clothing in Turkey using different econometric models indicating that the food expenditure is the largest expenditure and the portion of this expenditure in the household budget decreases as the income increases that is consistent with the Engel's law. Pritchett & Spivack (2013) introduced a new simple, intuitive appeal and consistent with the Engel's Law ratio measure for comparison of consumption possibilities over countries using average food shares and taking into account purchasing power of currencies and adjusted currency conversions. Gibson & Kim (2015) revealed that food shares vary with relative prices, but sometimes spatial price survey is not possible, and unit values are sometimes used as price proxy. Önder (2017) studied that the consumer consumption patterns can vary based on the geographical and socio-economic structure differences, the survey time and the specific country circumstances. Li (2021) empirically observed that richer consumers purchasing a larger variety of products than poor ones. Divergent Engel curve slopes rely on the relative price and transaction cost marginally and affect the distribution of variety gains and the measurement of factual welfare. Laborda et. al (2021) proposed the computation of total household expenditure in Household Budget Surveys (Engel curves) using a generalized linear model estimator, to deal with the heteroskedasticity problem encountered in the ordinary least squares method.

In our approach the following factors that affect Engel curve estimations were not taken into account:

- national characteristics (mentality, psychology, religious attitudes etc.),
- personal preferences and consumer behavior,
- different income groups behaviors,
- inflation, purchasing power and currency conversions,
- economical, geographical and regional differences.

Research objective of the paper is to forecast Engel curve components with the application of ARIMA time series forecasting method. The Engel curve evaluation is recurrently important to get information about average wellbeing situation in the country.

The paper is organized as following: introduction is followed by methodology that covers ARIMA process, then comes forecasting of Engel curve components and conclusions.

METHODOLOGY

A more important aspect of time series forecasting is whether it is stationary or not. In a broad sense, a stochastic process is considered stationary when its expected value and variance remain constant over time, and the covariance depends not on the time at which it is calculated, but on the difference between two consecutive times. Most stochastic processes are considered weakly stationary. Determining stationarity is important because if a time series is non-stationary, its study is only relevant at the current time. For this reason, non-stationary time series are brought to stationary series (integration or differencing) and forecasted with ARIMA (autoregressive integrated moving average) model. Thus, ARIMA is a genaralization of ARMA (autoregressive moving average) model. The model has both autoregression (AR) and moving average (MA) properties. The ARIMA(p', q) model is given by:

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_{p'} X_{t-p'} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_q \varepsilon_{t-q} \tag{1}$$

Or:

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t \tag{2}$$

Where, X_t – are forecast values at time t, L – is the lag operator, α_i – are autoregregression part parameters, θ_i – are moving average parameters, ε_t – is white noise, p – is the order of autoregression part built on its own lagged values, and q – is the order of moving average part respectively.

Now, let to assume that polynomial $\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right)$ has a unit root (a factor $(1 - L)$) raised to *d-th* power. Then it can be expressed as:

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) = \left(1 + \sum_{i=1}^{p'-d} \varphi_i L^i\right) (1 - L)^d \tag{3}$$

The general form of ARIMA(p,d,q) is expressed as below:

$$\left(1-\sum_{i=1}^{p}\varphi_i L^i\right)(1-L)^d X_t = \delta + \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t \qquad (4)$$

Where, the drift of ARIMA(p,d,q) proces is $\frac{\delta}{1-\sum\varphi_i}$, d – is the degree of differentiation.

ARIMA(p,d,q) process has the factor properties of polynomials with the difference $p = p' - d$. ARIMA process is a particular case of an ARMA(p+d, q) process which has the autoregressive polynomial with d unit roots. When nonstationary processes are brought into the ARIMA process they become stationary or weakly stationary.

ARIMA as a time series forecasting process is based on Box-Jenkins (2016) methodology. The method is built from four steps:

1. **Identification.** The adequate values of p, d and q are identified by using the correlogram and partial correlogram outputs.

2. **Calculation.** Next, the parameters p and q terms of the AR and MA are identified and included in the model using simple least squares method, but sometimes nonlinear (in parameter) calculation methods is also possible. For this purpose statistical tools (Eviews e.g.) are applied; the AR və MA parameters are obtained for each (p,d,q) set.

3. **Diagnostic checking.** The BJ methodology is applied as an iterative process for the selection of adequate ARIMA model. Having computed the parameters of the alternative models, the chosen one must be checked whether the residuals are white noise or not; if not, the process must be started over.

4. **Forecasting.** The ARIMA modeling is notable in forecasting due to its credibility compared to the econometric modeling, especially for short-term forecasts.

The Box-Jenkins methodology is the ground for selection of proper forecasting method among AR, MA, ARMA, and ARIMA (Gujarati, 2004).

FORECASTING OF ENGEL CURVE COMPONENTS

In this section, for the forecasting purposes household income and expenses per capita data are taken from State Statistical Committee of Azerbaijan (table 1). The data given in table 1 match the middle class of population in the country (stratified based on income per capita) and possibly represent the averaged Engel curve throughout the country.

Table 1: Household income and expenses per capita

№	Years	Income	Expenses	Expenses on food	Expenses on food (%)
1	2008	108.9	114.6	65.2	0.57
2	2009	125.0	129.6	68.6	0.53
3	2010	144.2	147.4	71.1	0.48
4	2011	166.0	173.0	82.4	0.48
5	2012	190.9	202.0	87.3	0.43
6	2013	214.7	221.4	91.8	0.41
7	2014	230.0	234.9	95.6	0.41
8	2015	240.5	245.6	99.4	0.40
9	2016	257.8	264.7	107.1	0.40
10	2017	268.4	278.2	117.9	0.42
11	2018	276.0	286.0	119.7	0.42
12	2019	292.6	298.4	123.8	0.41
13	2020	291.4	297.8	129.2	0.43
14	2021	300.6	308.6	134.7	0.44

First, the stationarity of the time series is checked according to the identification phase. Based on the correlogram and graphical analysis and ADF (Augmented Dickey Fuller) test, it is defined that whether the time series is non-stationary or has an increasing trend. In the given case the time series data are non-stationary, so based on correlogram outputs in Eviews program package the ARIMA models were constructed up to sixth differences.

Consequently the residuals are checked for White Noise using Ljung-Box Q Statistics; AR roots should lie inside a unit circle showing covariance stationarity and MA roots should lie inside a unit circle showing invertibility of ARMA process. Various models constructed with separate values of p, d, and q parameters are selected by the Akakike, Hannan-Quinn, and Schwarz statistics, whereby a smaller value identifies a better parameterized model.We tested the model estimation parameters with Q-statistics. Where the obtained probability values are greater than 0.05, the residuals are considered white noise. Based on computed covariances, stationarity of the processes is ensured. Below in tables 1-3 the ARIMA models are given, built on Box-Jenkins methodology.

Table 2: ARIMA model for household incomes

Dependent Variable: D(INCOME)				
Method: Least Squares				
Date: 01/08/23 Time: 10:57				
Sample (adjusted): 3 14				
Included observations: 12 after adjustments				
Convergence achieved after 17 iterations				
MA Backcast: 2				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.475553	13.10608	0.265186	0.7968
AR(1)	0.879029	0.128212	6.856054	0.0001
MA(1)	-0.999931	0.155679	-6.423015	0.0001
R-squared	0.589464	Mean dependent var		14.63333
Adjusted R-squared	0.498234	S.D. dependent var		7.604464
S.E. of regression	5.386656	Akaike info criterion		6.418045
Sum squared resid	261.1446	Schwarz criterion		6.539271
Log likelihood	-35.50827	Hannan-Quinn criter.		6.373162
F-statistic	6.461285	Durbin-Watson stat		2.438545
Prob(F-statistic)	0.018200			

Table 3: ARIMA model for household expenses

Dependent Variable: D(EXPENSES)				
Method: Least Squares				
Date: 01/08/23 Time: 12:58				
Sample (adjusted): 7 14				
Included observations: 8 after adjustments				
Failure to improve SSR after 13 iterations				
MA Backcast: 5 6				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.146122	21.55672	-0.053168	0.9609
AR(1)	0.095612	0.419764	0.227775	0.8345
AR(5)	0.489263	0.653582	0.748587	0.5085
MA(1)	-0.646581	1.235461	-0.523352	0.6369
MA(2)	0.999303	1.159314	0.861978	0.4521
R-squared	0.783570	Mean dependent var		10.90000
Adjusted R-squared	0.494997	S.D. dependent var		5.668459
S.E. of regression	4.028208	Akaike info criterion		5.893691
Sum squared resid	48.67938	Schwarz criterion		5.943342
Log likelihood	-18.57476	Hannan-Quinn criter.		5.558815
F-statistic	2.715327	Durbin-Watson stat		2.341917
Prob(F-statistic)	0.219031			

Table 4: ARIMA model for household expenses on food

Dependent Variable: D(EXPONFOOD)				
Method: Least Squares				
Date: 01/08/23 Time: 12:21				
Sample (adjusted): 8 14				
Included observations: 7 after adjustments				
Failure to improve SSR after 11 iterations				
MA Backcast: 2 7				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.787033	2.967827	2.286869	0.0841
AR(6)	0.549085	0.450273	1.219449	0.2897
MA(6)	-0.946182	0.342401	-2.763377	0.0507
R-squared	0.990023	Mean dependent var		5.585714
Adjusted R-squared	0.985034	S.D. dependent var		2.927700
S.E. of regression	0.358159	Akaike info criterion		1.081849
Sum squared resid	0.513112	Schwarz criterion		1.058668
Log likelihood	-0.786472	Hannan-Quinn criter.		0.795332
F-statistic	198.4574	Durbin-Watson stat		2.036730
Prob(F-statistic)	0.000100			

Using ARIMA models in tables 2-4 Engel curve components: household income, expenses, expenses on food were forecasted for the next five years (table 5). Expenses on food are calculated based on obtained data.

Table 5: Forecast results for Engel Curve components between 2022-2026

№	Years	Income	Expenses	Expenses on food	Expenses on food (% in total)
1	2022	309.10	315.76	141.71	0.45
2	2023	317.00	319.79	150.54	0.47
3	2024	324.36	325.76	154.98	0.48
4	2025	331.26	325.57	160.44	0.49
5	2026	337.73	330.36	166.44	0.50

According to table 5 it can be inferred that household income, expenses, and expenses on food forecast for 2022-2026 will show relatively gradual increment in comparison with factual data dinamics which markedly increased till 2021. It can also be derived that expenses on food will increase from 45% to 50%. The trend patterns of forecasted items are illustrated in figure 1.

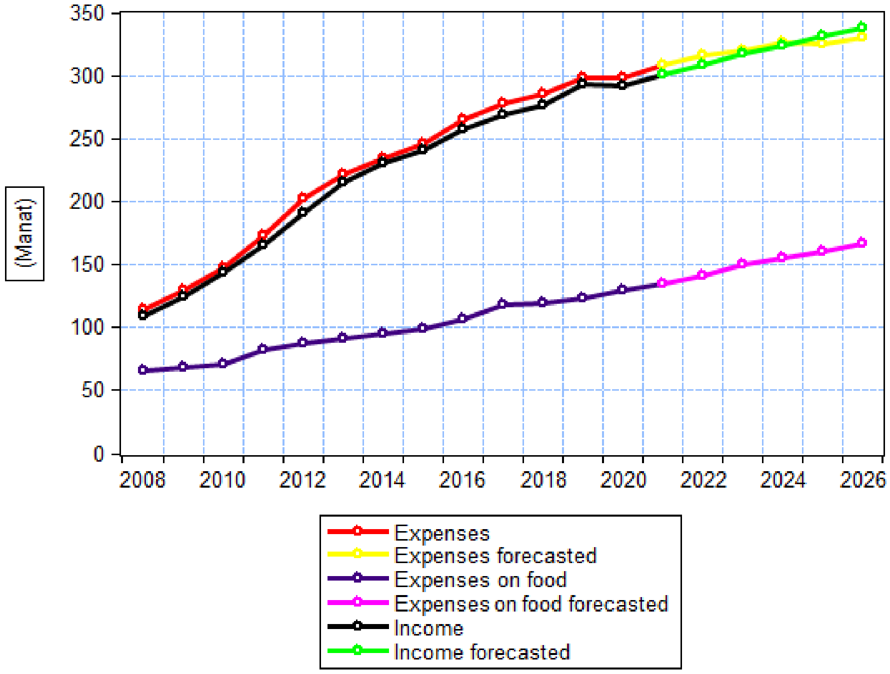

Figure 1: Forecast results of Engel Curve components between 2022-2026

Figure 1 information helps to infer that forecast results of household income, expenses and expenses on food will inrease in line with the trend lines. But, only household expences will go below the trend line.

CONCLUSION

In this paper Engel curve componenets are forecasted for 2022-2026 period with the application of ARIMA method. Engel curve as an express analytical and practical tool is effective to get some information about wellbeing and consumption patterns of households in the country. Despite that this methodology doesn't take into account consumption behaviour, changes in household income, changes in quality of consumption goods, mentalities and other factors, it is still applicable to get an overall picture about median household consumption structure on food. Since Engel curve is not applied to all income groups, the forecasted data correspond only to median households in the country. ARIMA time series forecasting model is one of the widely used econometric methods in this regard to predict average change. The obtained results can be useful for decision-making in socio-economic sectors.

REFERENCES

Almas, Ingvild. (2012), “International Income Inequality: Measuring PPP Bias by Estimating Engel Curves for Food.” American Economic Review, 102(2): 1093-1117.

Banks, J., Blundell, R., Lewbel, A., (1997), Quadratic Engel Curves and Consumer Demand, The Review of Economics and Statistics, 79(4): 527-539.

Blundell, R., Duncan, A., Pendakur, K., (1998), Semiparametric Estimation of Consumer Demand, Journal of Applied Econometrics 13(5): 435-461.

Box, G., Jenkins G., Reinsel G., Ljung G. (2016), Time Series Analysis: Forecasting and Control, New Jersey: Wiley, p.709.

Çağlayan, E. (2012), An econometrıc analysıs of engel’s curve: household food and clothıng consumptıon ın Turkey, Scientific Annals of the “Alexandru Ioan Cuza” University of Iaşi Economic Sciences, 59 (1), 313-319.

Engel, E. (1857), “Die Productions- und Consumtionsverhältnisse desKönigreichs Sachsen,” Zeitschrift des statistischen Bureaus des Königlich Sächsischen Ministerium des Inneren 8–9: 28–29.

Gibson, J., Kim B. (2015), Hicksian Separability Does Not Hold Over Space: Implications for the Design of Household Surveys and Price Questionnaires. Journal of Development Economics 114 (1): 34–40.

Gujarati, D.N. (2004), Basic Econometrics, NY: The McGraw Hill Companies, 1003 p.

James, L. Seale Jr, Junfei, Bai, Thomas, I. Wahl, Bryan, T. Lohmar, (2012),"Household Engel curve analysis for food, Beijing, China", China Agricultural Economic Review, 4(4), 427 – 439.

Kumar, T. K., Holla, J., Guha, P. (2008), Engel curve method for measuring poverty, Economics and Social Science Area Indian Institute of Management, Conference of the Indian Society of Probability and Statistics, Nagpur, India.

Leive, A., Xu, K. (2007), Coping with out-of-pocket Health Payments: Applications of Engel Curves and Two-Part Models in Six African Countries, Discussion Paper, World Health Organization, Geneva.

Li, N. (2010), An Engel Curve for Variety, The Review of Economics and Statistics, MIT Press, 103(1), p. 72-87.

Lopez-Laborda, J., Marin-Gonzales C., Onrubia-Fernandez, J. (2021), Estimating Engel curves: a new way to improve the SILC-HBS matching process using GLM methods, Journal of applied statistics, 48(16), 3233–3250.

Önder, K. (2017), The Effect of the Period that Turkey is going Through onthe Household Expenditures of Burdur Province: EngelCurve Analysis, Turkish Economic Review, 4(2), 216-225.

Pritchett, L., Spivack, M. (2013), Estimating Income/Expenditure Differences across Populations: New Fun with Old Engel's Law, Working paper 339, Center for Global Development.

Solomon, Michael, Gary B., Soren A., and Margaret H. (2006), Consumer Behavior: A European Perspective, Third Edition, Harlow, Essex: Prentice Hall.

You, J. (2003), Robust Estimation of Models of Engel Curves, Empirical Economics, 28(1): 61-73.