JOURNAL OF APPLIED MICROECONOMETRICS

HOLISTENCE publications

E-ISSN: 2791-7401
DOI: 10.53753/jame

International Peer-Reviewed and Open Access Electronic Journal

JAME
JOURNAL OF APPLIED MICROECONOMETRICS

Volume: 3
Issue: 1
2023

E-ISSN: 2791-7401
DOI: 10.53753/jame

International Peer-Reviewed and Open Access Electronic Journal

Volume: 3

Issue: 1

Year: 2023

WEB: https://journals.gen.tr/jame

E-mail: jame.editors@gmail.com

Address: Sarıcaeli Köyü ÇOMÜ Sarıcaeli Yerleşkesi No:29 D:119
Merkez-Çanakkale/TÜRKİYE

ABOUT THE JOURNAL

The Journal of Applied Microeconometrics (JAME) is an internationally peer reviewed and periodical open access e-journal covering any issues in theoretical and applied microeconometrics. The journal also covers quantitative research in microeconomics. The main topics subject areas of the journal are as follows:

- Estimation and Testing
- Robust Regression
- Discrete Choice Modelling
- Models for Count Data
- Duration Models
- Limited Dependent Variables
- Microeconometric Theory
- Bayesian Microeconometrics
- Big Data and Machine Learning in
- Microeconometrics
- Spatial Econometrics
- Non-Parametric Regression

JAME aims to serve as a platform for high quality research in applied microeconometrics. The scope of the Journal includes any papers dealing with identification, modelling, estimation, testing and prediction issues encountered in the analysis of individual-level data on the economic behavior of individuals or firms using regression methods for microeconometric data i.e. cross sectional data, repeated cross- sectional data, pool data, cohort and panel data etc. The journal also accepts case study articles written for both developing and developed countries.

We receive seven types of submissions:

1- Regular submissions (open submissions)

2- A special issue

a) consisting of selected full length articles from national/international conferences, symposiums or workshops on microeconometrics.

b) by a guest editor invited by the editor in chief focusing on a specific area of research in microeconometrics.

3-Critiques of Previously Published Research Papers

4-Case Study Articles

5-Book Reviews

6-Letters to the Editor

7-Invited articles (Invited articles are submissions of which authors are invited to publish their work by the editor in chief and not subject to a review process.)

The publication language of the journal is English.The Journal of Applied Microeconometrics is an open access electronic journal. Each paper published in the Journal is assigned a DOI® number, The journal is published 2 times a year as SUMMER (June) and WINTER (December) periods.

Send your manuscript to the editor at : https://journals.gen.tr/jame/user/register/

Owner

Holistence Publications

Contact

Adress: Sarıcaeli Köyü ÇOMÜ Sarıcaeli Yerleşkesi No:29, D.119, Merkez-Çanakkale / TÜRKİYE

Tel: +90 530 638 7017

WEB : https://journals.gen.tr/jame

E-mail : jame.editors@gmail.com
GSM 2: +90 530 638-7017 / WhatsApp

EDITORS

CONTENTS

"This page is left blank for typesetting"

JAME
E-ISSN:2791-7401
Volume : 3, Issue 1, 2023
URL: https://journals.gen.tr/jame

Received: 2023-07-03
Accepted: 2023-07-05

INVITED ARTICLE

A semi-nonparametric extended ordered probit model with selection for financial barrier perception

Hülya ÜNLÜ

Department of Economics / Cankiri Karatekin University, Türkiye, e-mail: hunlu@karatekin.edu.tr

Abstract

In order to contribute to the literature concerning the difficulties faced by innovative firms in terms of financing, this paper aims to investigate the perception levels of financial barriers according to their innovation intensity and analyzes determinants of financial barriers for a developing country for the most recent years. A semi-nonparametric extended ordered probit model with selection is used to establish the determinants of perception of financial barriers by employing the Business Enterprise and Environment Survey, BEEPS 2013 and BEEPS 2019. According to the findings, when there is an engagement in innovation activities, then firms are more likely to assess financial barriers as important. It is believed that these results have important implications for developing countries.

Keywords: Financial Barriers, Innovation Investments, Revealed Barriers, Semi-nonparametric, Selection Bias

JEL Codes: O30, G32, L00, C25

Citation: ÜNLÜ, H. (2023). A semi-nonparametric extended ordered probit model with selection for financial barrier perception. *Journal of Applied Microeconometrics (JAME)*. 3(1), 1-10, DOI: 10.53753/jame.2097

Corresponding Author:
Hülya ÜNLÜ
E-mail: hunlu@karatekin.edu.tr

1. INTRODUCTION

Changing human needs keep two main issues on the agenda. The first of these is the tendency to innovate, which is the main focus of today's entrepreneurs in order to catch up with changing human needs; while the second one is to find the necessary financing to realize the first. The newly created markets brought about by creative destruction and the desire to quickly respond to people's needs through these markets are supported by the profit motive. Today, it is known that companies that would like to maintain their profit advantage, especially after the 2000s, aim to evaluate all their opportunities for innovative activities. However, fund owners who are aware that innovative processes are full of uncertainties are more reluctant to invest their funds in these areas if they are not risk-takers. For this reason, entrepreneurs who see the hesitations of finance owners have the perception that they have difficulties in accessing finance. Even though many developed countries have more financing opportunities for risky investments such as angel investors and venture capital. For a developing country like Türkiye, it is possible to say that the financial markets are not yet fully ready for innovative product and service investments where uncertainty is intense (see detailed literature for Türkiye; (Ünlü 2022; Ünlü, Çankal and Çetin 2022)).

Many researchers have been exploring the cause of low levels of innovation across countries by emphasizing that successful innovation processes depend on important skills (D'Este et al. 2012). These skills also affect the innovation intentions of the companies (Iammarino, Sanna-Randaccio and Savona 2009; D'Este et al. 2012; Almeida, Hsu and Li 2013; Guariglia and Liu 2014). For this reason, the literature focuses on outlining the factors that determine the barriers to innovation (Tiwari et al. 2007; Canepa and Stoneman 2008; Mohnen et al. 2008). Developing countries are particularly interested in new policy frameworks related to Science, Technology, and Innovation (Santiago et al. 2017). As mentioned before, Türkiye is a country with high potential and has directed its policies towards innovation, especially in technology development regions, in the last 20 years. For this reason, it is important to determine the problems in the system from different perspectives in order for the policies to be effective at the desired level. Considering all these, in this study, the relationship between the level of perception of financial barriers and their tendency to innovate was examined in terms of enterprises. Differently from the common literature in this study, the possible selection bias is considered. Our main question is "How does the perception of financial barriers of enterprises change according to whether the enterprises tend to innovate or not". In the study, company characteristics, sectoral differences and regional differences were taken into account. Since the obstacle perception used in the study was measured with a five-point Likert scale, using the ordered probit model is appropriate. However, since it is known that the semi-nonparametric ordered probit model (SNPOPM) relaxes the normal distribution assumption, the use of SNPOPM has been preferred. Savignac (2008) suggests that there may be a selection bias towards innovative firms and non-innovative firms. While the literature defines innovative companies differently according to the survey used, our survey does not include any questions about whether the company is willing to innovate, so it is not possible to follow Savignac (2008). The reason why we follow the same structure as Männasoo and Meriküll (2020) is that we use similar survey data. To control for selection bias, we used a semi-nonparametric extended ordered probit model (SNPEOPM) with selection. Differently from the existing literature with the best knowledge of the author, this paper is the first paper that takes into account several problems such as heteroscedasticity, normality and sample bias. The analysis concluded that Turkish businesses perceive financial barriers as high when they are innovatively active, as suggested in previous literature (D'Este et al. 2012; Santiago et al. 2017; de-Oliveira and Rodil-Marzábal 2019; Ünlü, Çankal and Çetin 2022). We consider a selection model and the result of the selection model suggests that to perceive the barriers as less important firms should be aged and bigger sized.

2. DATA AND METHOD

The data is derived from the most recent Business Environment and Enterprise Performance Survey (BEEPS) 2013 and BEEPS 2019 surveys, which include innovation activities and financial barriers (for more information: World Bank Enterprise Surveys, http://www.enterprisesurveys.org). The sample consists of 1344 firms for 2013 and 1663 firms for 2019. BEEPS data collect a broad set of information about innovation, perceiving access to finance as a barrier, firm characteristics, and the business environment. In the data, because only a few firms are visited again, similar to Männasoo & Meriküll (2020), it is preferred to use pooled cross-section. After cleaning the data from nonresponses, the new sample size for all 2730. The financial barrier used in the study is on a 5-point Likert scale in response to the question of "How Much of An Obstacle: Access To Finance?" (from no obstacle to very severe obstacle). The strength of this database is that it measures financial barriers directly and provides a clear indication when compared to indirect measurement based on cash flow (Männasoo and Meriküll 2020). Thus, it is possible to reach information on whether companies have problems in accessing loans.

However, by examining the distribution of financial barrier perception, whose model preference is an ordered variable, it was found appropriate to use one of the ordered type probit models. Because the ordered probit model (OPM) has the assumption that the error term should be normally distributed (Wooldridge 2013), it is necessary to test whether the appropriate model is the OPM or not. We use a semi-nonparametric (SNP) extended ordered probit model which relaxes the normality assumption (Stewart 2004)[1].

Stewart (2004) developed the SNP approach from the suggestion of that for a consistent SNP estimator, the unknown density must be sufficiently smooth with an upper bound in the tails. Just as they suggest that the estimator can accommodate density skewness and kurtosis and fail only when the density is strongly oscillating, Stewart (2004) normalizes the model using the estimation from an ordered probit for the first cutoff point. According to Stewart (2004), the SNP approach is using a pseudo maximum likelihood estimator for the vector of model parameters, to be able to do this the SNP approach estimates the unknown densities of the error terms by Hermite polynomial expansions. Different from the ordered probit model, the SNP extended probit model's interpretations depend on the K selection, where the K shows the number of values given for an ordered variable (Vieira at al. 2023). It is known that the SNP takes the K=3 as the lowest possible value and the system for K<3, the model crashes to the ordered probit state. For this reason, a model selection depends on the K and can be justified using likelihood-ratio tests or the Akaike information criterion (Doremus 2020).

Before going further in this analysis we also checked possible heteroscedasticity by using the heteroscedastic OPM (Keele and Park 2006). In this study, both industry and service sectors are included, as also applied in the study of D'Este et al. (2012). However, to ensure homogeneity, sector variables and region variables are added to the model as dummy variables. As Savignac (2008) proposes that there might be a selection bias towards innovative firms and non-innovative firms. While literature defines innovative firms differently concerning the survey used, our survey does not include any question regarding whether the firm does not innovative because of willingness, it is not possible to follow Savignac (2008). We followed the same structure with Männasoo and Meriküll (2020) and we also checked whether thereis a selection bias or not. To control selection bias, we used a semi-nonparametric extended ordered probit model (SNPEOPM) with selection[2]. To test this issue, it is used SNPEOPM with selection, which is proposed by De Luca and Perotti (2011).

Table A1 (Appendix) demonstrates the detailed information about the explanatory variables and provides information related to the descriptive statistics. The engagement level of innovation activities is determined from three main questions related to research and development activities; "*During last 3 years, establishment spent on the acquisition of external knowledge?", "… on R&D within the establishment?",* "*… on R&D contracted outside establishment?".* The innovation engagement has determined as if the firm answered yes to at least one of these questions. The Chi2 test is provided in Table A2 (Appendix) for the companies identifying obstacles as important based on the level of innovation participation. The outcome provides a statistically significant test statistic for the hypothesis that the degree of a firm's innovation activity is not independent of the evaluation of obstacles. It demonstrates that companies with low involvement levels are more likely than companies with high participation levels to believe that financial obstacles to innovation are substantially less relevant..

3. RESULTS

The results in this study follow several steps: first, we aimed to prove that there was not any heteroscedasticity problem. With the aim to reveail this, we used heteroscedastic ordered probit[3] model which provides a likelihood ratio test of homoscedasticity. First, the age of firms suggests that the older firms might have the experience to deal with risky funding of innovation than the younger ones. Thus, the variance of the error terms of age may influence the variance of unobserved heterogeneity. Second, the size of the firms may also influence the variance of the unobserved heterogeneity. The results are summarized in Table A3 (Appendix) LR test gives a chi2 of 0 when the age of the firms given as a source of heteroscedasticity and a similar result seen when the source is given as employment with the chi2 of 0.14. The result shows that there is no heteroscedasticity problem for the given models.

The second step of the analysis is to investigate whether the assumption of the normality for the OPM was violated or not. Table 1 shows the OPM and SNPEOPM estimator results. It is seen from Table 1 that the likelihood ratio (LR) test of the OPM against the SNPEOPM (where K=5) gives Chi2 statistic equal to 144 and rejects the null that the OPM is more suitable than the SNPEOPM. The AIC is lowest for the case of K = 5. A LR test (4.31***) comparing K = 5 to K = 4 can reject that K = 4 fits the data better. Because of the lower AIC and rejection of the hypothesis that K = 4 is a better fit, therefore, the results suggest to select the K = 5 model.

Table 1. Likelihood Ratio Test of OPM and SNPEOPM

K	*Log-likelihood*	*LR test of ordered probit*	*Degrees of freedom*	*LR test of K-1*	*AIC*
OP	-3454.78				6959
4	-3455.86	140***	2	-2.16	6963
5	-3453.71	144***	3	4.31***	6961

Note: *** p<0.01, ** p<0.05, * p<0.1

To show whether there is a selection bias or not, we have used the special model which is developed for sample selection for the SNPEOPM. While the literature proposes that the innovators behave differently than non-innovators against the financial obstacles, we proposed the model that the innovators who claim product and process innovation has been done by their company should be used as a sample selection model. As De Luca and Perotti (2011) suggest that any unobservable factors that are probable might affect the outcome of interest, which may cause inconsistent estimates of the SNPEOPM. We explained the innovators by three instruments given in the literature. According to Männasoo and Meriküll (2020), competitive features can be used, unlike the classic firm characteristics. This feature can be explained by whether firms see anti-competitive practices as a serious obstacle to the business, and obstacles to qualified personnel and training opportunities can also be used as an indicator for innovators. The results of the alternative SNPEOPM with the selection model are given in Table 2. Both LR tests and AIC are given in the table and suggest that the model with order (5,5) is more appropriate.

Table 2. Likelihood Ratio Test of SNPEOPM with Selection

K	*Log-likelihood*	*LR test of K-1*	*AIC*	*Rho*
(3,3)	-6541.81		13151	-0.18
(4,4)	-6505.27	73.07***	13092	-0.50
(5,5)	-6495.69	19.16***	13091	-0.59

Note: *** p<0.01, ** p<0.05, * p<0.1

The result of the SNPEOPM with selection is given in Table 3 and suggests that there is a negative correlation between the error terms of the main model and selection model, which is -0.59. This indicates selection bias. However, the significant variables are still significant. And the signs of the estimators do not change. According to Table 3, it is found that if the innovation activity engagement is high then there is a significant and a positive effect on the perception of financial barriers, which suggests revealed effect as suggested by D'Este et al. (2012).

Table 3. Estimation results for OPM, SNPEOPM (K=5), and SNPEOPM with Selection (R1=5, R2=5)

VARIABLES	OPM	SNPEOPM	SNPEOPM with Selection
Engagement To Innovation Activities	0.21***	0.12***	0.12***
	(0.057)	(0.049)	(0.045)
Age	-0.045	-0.046***	-0.103***
	(0.029)	(0.023)	(0.021)
Size	-0.067***	-0.029**	-0.047***
	(0.018)	(0.015)	(0.015)
Learning by Export	-0.000	-0.026	0.023
	(0.055)	(0.044)	(0.045)
Group Engagement	0.189***	0.114***	0.145***
	(0.068)	(0.055)	(0.053)
Sector dummies	Included	Included	Included
Regional dummies	Included	Included	Included
Year 2019	Included	Included	Included
Sample selection model			
Competition as an obstacle			0.370***
			(0.09)
Uneducated workers as an obstacle			0.418***
			(0.125)
Training for employees			0.219***
			(0.092)
Intercept			-1.241
Cut 1	-0.274		
Cut 2	0.264		
Cut 3	1.270		
Cut 4	2.032		
Thresholds 1		-0.27	0.64
		(Fixed)	(Fixed)
2		0.14	1.09
		(0.07)	(0.03)
3		0.96	1.95
		(0.21)	(0.09)
4		2.13	3.00
		(0.43)	(0.15)
SNP Coefs.		estimated	estimated
LR Chi2	683***		
Wald Chi2		33.55**	
Pseudo R2	0.088		
OBSERVATIONS	2730	2730	2570
Estimated moments of errors distribution		Main Equation	Selection Equation
Standard Deviation		1.457927	2.099603
Variance		2.125552	4.408331
Skewness		-1.080913	-.4749198
Kurtosis		4.668988	2.24423

Note: a) Standard errors are given in parentheses. b) *** p<0.01, ** p<0.05, * p<0.1. c) **OPM:** ordered probit model, **SNPEOPM:** semi-nonparametric extended ordered probit model and **SNPEOPM** with selection: semi-nonparametric extended ordered probit model with selection, respectively.

Table 4. Hypothesis Testing

Null Hypothesis	LRT	DF	p-value
Engagement to Innovation Activities has no influence on perception of financial barriers, all else equal	46.21	10	0.000

Note: LRT, DF and p-value represent likelihood ratio tests, degrees of freedom (equivalent to the number of constraints imposed on the model) and test p-values (for a chi square distribution).

This result is also consistent with our hypothesis (Table 4) that innovative-active firms are more likely to face financial barriers to innovation and therefore more likely to perceive financial barriers as significantly higher. As the literature suggests, firm size significantly affects the perception of barriers to innovation. More specifically, larger companies perceive financial barriers as less relevant than smaller companies. The case of older / mature firms similarly perceives financial barriers less using their experience. While "learning by export" is statistically nonsignificant, it is seen that firms in a group have an increasing effect on the perception of financial barriers contrary to what is expected. It is seen that the impact differs according to the sectors and regions. The sectoral divergence is also seen from the study, that the garments and other manufacturing firms are more likely to face financial barriers rather than service sectors. The selection model has significant explanatory variables.

4. CONCLUSION

In this study, it was investigated which companies felt the financial barriers. In particular, the perception of companies engaged in intensive innovation activities was examined. In the study, the BEEPS data was used for Türkiye, and the years 2013 and 2019 were included. Although barriers have been examined in various studies before, models measuring perception at different levels are less common in the literature. For this reason, the use of the ordered probit model was preferred in the study. Sample selection model is included to address this issue as selection bias can be an issue. In addition, as a result of the control of the assumptions, the use of semi-non parametric model was found more appropriate. The study differs from other studies in terms of method. When the results are examined, as expected, age and size, which are the main characteristics of the company, affect the perception of financial barrier. Beck et al. (2006) supports our finding that age and size have negative effect,which means that younger and smaller firms are feeling major financial obstacles. While the existence of entrepreneurs being a member of a business group creates a significant effect, firms that export unexpectedly do not have any significant effect. Especially in the study, the result of the significant effect of companies with innovation activities that carry out risky activities is striking. According to the literature, it was expected that firms belonging to a business group would face lower financing barriers as they have access to the group's internal cash flow, but the positive and significant sign obtained is that if Turkish firms are included in the group, they tend to have difficulties to diverge in income and expenses both in the same business line or in different business lines. For this reason, if a member has financial difficulties, it may be due to the fact that other members have similar problems. This draws our attention as an issue that needs to be studied in the future.

Acknowledgment

The authors extend sincere gratitude to the Editors who provided very constructive feedback for revision, improvement, extension, and finetuning of the paper for publication. We thank the Enterprise Analysis Unit of the Development Economics Global Indicators Department of the World Bank Group for making the data available.

Funding

This research did not receive any specific grant from the public, commercial, or not-for-profit funding agencies.

Conflict of Interest

All authors declare that there is no conflict of interest.

Submission Declaration Statement

We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere.

Endnotes

[1] More detail about the SNP extended probit model can be seen from the Stewart (2004).

[2] More detail about the SNP extended probit model with selection can be seen from the De Luca and Perotti (2011).

[3] The estimates are done by using stata command hetoprobit.

REFERENCES

- ALMEIDA, H., HSU, P.-H. AND LI, D. (2013) Less is More:Financial Constraints and Innovative Efficiency. *SSRN 1831786*. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1831786.
- BECK, T., DEMIRGUÇ-KUNT, A., LAEVEN, L., & MAKSIMOVIC, V. (2006). The Determinants of Financing Obstacles. *Journal of International Money and Finance*, 25(6), 932-952.
- CANEPA, A. and STONEMAN, P. (2008). Financial Constraints to Innovation in the UK: Evidence from CIS2 and CIS3. *Oxford Economic Papers*, 60(4), 711–730. doi: 10.1093/oep/gpm044.
- D'ESTE, P., IAMMARINO, S., SAVONA, M., & VON TUNZELMANN, N. (2012). What Hampers Innovation? Revealed Barriers Versus Deterring Barriers. *Research Policy*, 41(2), 482-488.
- DE-OLIVEIRA, F. and RODIL-MARZÁBAL, Ó. (2019). Structural Characteristics And Organizational Determinants as Obstacles to Innovation in Small Developing Countries. *Technological Forecasting and Social Change*, Elsevier, 140, 306–314. doi: https://doi.org/10.1016/j.techfore.2018.12.021.
- DE LUCA, G. and PEROTTI, V. (2011)/ Estimation of Ordered Response Models with Sample Selection. *The Stata Journal: Promoting Communications on Statistics and Stata*, 11(2), 213–239. doi: 10.1177/1536867X1101100204.
- DOREMUS, J. (2020). How Does Eco-Label Competition Affect Environmental Benefits? The Case Of Central Africa's Forests. *Journal of Environmental Economics and Management*, 103, 102344. doi: 10.1016/j.jeem.2020.102344.
- GUARIGLIA, A. and LIU, P. (2014). To What Extent Do Financing Constraints Affect Chinese Firms' Innovation Activities?. *International Review of Financial Analysis*. Elsevier Inc., 36, 223–240. doi: 10.1016/j.irfa.2014.01.005.
- IAMMARINO, S., SANNA-RANDACCIO, F. and SAVONA, M. (2009). The Perception Of Obstacles To Innovation. Foreign Multinationals and Domestic Firms in Italy. *Revue D ' Économie Industrielle*, 125. Available at: http://rei.revues.org/3953.
- KEELE, L. and PARK, D. K. (2006). Difficult Choices: An Evaluation of Heterogeneous Choice Models in *Paper For The 2004 Meeting of The American Political Science Association*, 2–5.
- MÄNNASOO, K. and MERIKULL, J. (2020). Credit Constraints And R&D over the Boom and Bust: Firm-Level Evidence From Central and Eastern Europe. *Economic Systems*. Elsevier, 44(2), 100747. doi: 10.1016/j.ecosys.2020.100747.
- MOHNEN, P., PALM, F. C., VAN DER LOEFF, S. S., & TIWARI, A. (2008). Financial Constraints and Other obstacles: Are They a Threat to Innovation Activity?. *De Economist*, 156(2), 201–214. doi: 10.1007/s10645-008-9089-y.
- SANTIAGO, F., DE FUENTES, C., DUTRÉNIT, G., & GRAS, N. (2017). What Hinders Innovation Performance of Services and Manufacturing firms in Mexico ?. *Economics of Innovation and New Technology*. Taylor & Francis, 26(3), 247–268. doi: 10.1080/10438599.2016.1181297.
- SAVIGNAC, F. (2008). Impact of Financial Constraints on Innovation: What Can Be Learned from A Direct Measure?. *Economics of Innovation and New Technology*, 17(6), 553–569. doi: 10.1080/10438590701538432.
- STEWART, M. B. (2004). Semi-nonparametric Estimation of Extended Ordered Probit Models', (1), 27–39.
- TIWARI, A., MOHNEN, P., PALM, F., & SCHIM VAN DER LOEFF, S. (2007). Financial Constraint and R&D Investment: Evidence from CIS. *Working Paper Series MERIT*, (31), 1–45.
- UNLU, H. (2022). The Effect of Credit Constraints on Innovation Activities During Different Economic Conditions: Firm-Level Evidence From Turkey. *International Journal of Business Innovation and Research*. Inderscience Publishers, 29(2), 221–234. doi: 10.1504/IJBIR.2022.126030.
- UNLU, H., ÇANKAL, E. and ÇETIN, A. K. (2022). Financial Constraints to Innovative Activities Revealed Barriers Versus Deterring Barriers: Evidence from Turkey, in Leitão, J. and Ratten, V. (eds) *Strategic Innovation: Research Perspectives on Entrepreneurship and Resilience*. Cham: Springer International Publishing, pp. 47–75. doi: 10.1007/978-3-030-87112-3_5.
- Vieira, J. A. C. et al. (2023) Climbing the ladders of job satisfaction and employee organizational commitment: cross-country evidence using a semi-nonparametric approach. *Journal of Applied Economics.* Routledge, 26(1). doi: 10.1080/15140326.2022.2163581.
- WOOLDRIDGE, J. M. (2013) *Introductory Econometrics: A Modern Approach*. Cengage Learning. Available at: https://books.google.com.tr/books?id=C0KHwUKxys0C.

APPENDIX

Table A1. Description of Variables and Descriptive Statistics

Type	Variable	Description	N	Overall Sample Mean	Std. Dev.	Min.	Max.
Dependent Variables	Financial Obstacles	=1 No Obstacle =2 Minor Obstacle =3 Moderate Obstacle =4 Major Obstacle =5 Very Severe Obstacle		2.27	1.19	1	5
Independent Variables							
Engagement to Innovation Activities	Innovation Active	=1 if the firm has at least one innovation activity =0 otherwise	2730	0.19	0.39		
Learning by Export	Export	= 1 if the share of direct and indirect exports in firm sales higher than 50% =0 otherwise		0.28	0.45		
Group Engagement	Part of Large Group	=1 if the firm is the part of Large Group =0 otherwise		0.12	0.32		
Innovative Firms		=1 if the firm has introduced new or significantly improved products or Services, or introduced new or significantly improved process, =0 otherwise		0.19	0.39	0	1
Compete		=1 if the anticompetitive practices are a serious concern for the firm, =0 otherwise	2628	0.26	0.44		
Uneducated workforce		=1 if the firm perceives Inadequately educated workforce as "Major obstacle" or "Very severe obstacle", =0 otherwise	2697	0.21	0.40		
Training		=1 if Formal Training Programs For Permanent, Full-Time Employees is given inn Last FY =0 otherwise	2694	0.35	0.47		
Firm Characteristics	Age of the Firm	Log(age)		2.67	0.77	0	4.59
	Size of the Firm	Log (employee)		3.40	1.37	0	8.55
Sector (S) 1	Food			0.12	0.32		
S2	Textiles		2730	0.12	0.32		
S3	Garments			0.11	0.31		
S4	Fabricated Metal Products	Dummy variable		0.11	0.31	0	1
S5	Machinery and Equipment			0.05	0.23		
S7	Construction			0.06	0.25		
S8	Retail			0.10	0.30		
S9	Other Services			0.08	0.28		
Region (R) 1	Marmara			0.27	0.44		
R2	Aegean			0.14	0.34		
R3	Mediterranean			0.16	0.36		
R4	Central Anatolia			0.11	0.32		
R5	Black Sea			013	0.33		
R6	Eastern and Southeastern Anatolia			0.17	0.38		
Year 2013				0.44	0.49		
Year 2019				0.55	0.49		

Table A2. The Percentage Of Firms Reporting Barriers As Important By Degree Of Engagement In Innovative Activities

	Engagement to innovation activities	
Financial Barriers	No	Yes
No Obstacle	38.76	32.10
Minor Obstacle	15.59	24.54
Moderate Obstacle	29.57	27.68
Major Obstacle	12.25	9.96
Very Severe Obstacle	3.84	5.72

Note: Pearson Chi2=31.29 and p=0.000

Table A3. Heteroskedastic OPM Results

	Financial Obstacles	**Financial Obstacles**
VARIABLES	Age as a source of heteroskedasticity	Size as a source of heteroskedasticty
Engagement To Innovation Activities	0.21***	0.22***
	(0.059)	(0.060)
Age	-0.044	-0.046
	(0.030)	(0.030)
Size	-0.06***	-0.07***
	(0.019)	(0.021)
Learning by Export	-0.000	0.000
	(0.055)	(0.057)
Group Engagement	0.18***	0.19***
	(0.069)	(0.071)
Sector dummies	Included	Included
Regional dummies	Included	Included
Year 2019	Included	Included
Lnsigma	Age	Size
	-0.001	-0.008
	(0.025)	(0.014)
Cut 1	-0.273	-0.279
Cut 2	0.264	0.264
Cut 3	1.267	1.300
Cut 4	2.027	2.083
LR test of lnsigma=0	0.00	0.35
Chi2	Pr=0.96	Pr=0.55
Wald test of lnsigma=0	0.00	0.35
Chi2	Pr=0.96	Pr=0.55
OBSERVATIONS	2730	2730

JAME
E-ISSN:2791-7401
Volume : 3, Issue 1, 2023
URL: https://journals.gen.tr/jame

Received: 2022-12-26

Accepted: 2023-03-20

RESEARCH ARTICLE

Determinants of household savings rates: Logistic quantile regression approach

Şaban Kızılarslan[1] **Serdar Göcen**[2]

[1] Corresponding author, Department of Econometrics, Faculty of Economics and Administrative Sciences, Van Yüzüncü Yıl University, Türkiye, e-mail: saban.kizilarslan@marmara.edu.tr

[2] Department of Economics, Faculty of Economics and Administrative Sciences, Osmaniye Korkut Ata University, Türkiye, e-mail: serdargocen@osmaniye.edu.tr,

Abstract

Saving rates have a fundamental economic importance that affects the economic performance of countries and the welfare level of individuals. Savings have been addressed in various ways with alternative economic approaches. The determinants of household savings rates were examined at the level of quantiles in this study. For this purpose, the logistic quantile regression approach proposed for bounded dependent variables was used. Since savings rates have a bounded and continuous structure, it is appropriate to analyze them with this method. Income level has been considered the principal determinant of savings rates and the change in the effect of income on savings rate at the level of quantiles was examined in details. As a result of the analysis performed separately for homeowners and tenants, it was determined that there were differences between the two groups. The change in the income effect was non-linear at the quantile level in both groups. While income was more effective at high savings rates for homeowners, it was more effective at low savings rates for tenants. On the other hand, the effects of other characteristics of the households also differed between the homeowners and the tenants at the level of the quantiles.

Keywords: Saving Rates, Income, Household, Logistic Quantile Regression, Bounded Outcomes

JEL Codes: C31, D14

Citation: KIZILARSLAN, Ş. & GÖCEN, S. (2023). Determinants of household savings rates: Logistic quantile regression approach. *Journal of Applied Microeconometrics (JAME)*. 3(1), 11-22, DOI: 10.53753/jame.1944

Corresponding Author:
Şaban Kızılarslan
E-mail: saban.kizilarslan@marmara.edu.tr

1. INTRODUCTION

Savings, which consist of the non-consumed part of the disposable income, have been handled in different ways by the schools of economics. According to the classical school of economics, consumption and saving decisions are determined by interest rate. Economic actors divide their income between consumption and savings according to the current interest rate. Savings, which have a positive relationship with interest rates, turn into investments, increasing the capital stocks and growth performances of countries. In Keynesian economic thought, this positive approach towards savings has changed. The situation which is called the saving paradox by Keynes (1936) is defined such that the increase in savings causes a decrease in production and employment by reducing demand and consumption expenditures without causing an increase in investments. Thus, a skeptical approach to saving was developed by Keynes, and the view that saving is harmful to the economy emerged (Modigliani 1986: 704).

According to the growth theories, the level of savings is vital for the economic performance of countries. In economic growth theories, in addition to exogenous theories such as Solow (1956) model and the Harrod-Domar model (Harrod 1939; Domar 1946), savings have positive effects on growth in endogenous growth theories (Lucas 1988). In addition, savings are decisive in explaining the differences in economic development between countries. One of the important factors in countries having different levels of development is the difference between capital stocks. Countries with low capital stocks lag in economic welfare in the long run, due to their slower growth. For a developing country like Türkiye, one of the key factors in achieving the welfare levels of developed countries is to increase capital accumulation. Savings are the element that allows for the necessary investments to expand capital stock.

Another factor in which savings are important is its relationship with the current account deficit, which is particularly important for the Turkish economy (Rijckeghem and Üçer 2009: 14). The main issue with countries having high current account deficits is low savings inspite of high investments. Low savings lead to the use of foreign resources in order to finance growth along with the increase in capital stock which results in a current account deficit. Hence, it is crucial to increase savings to reduce the current account deficit.

The first approach to explain saving behavior is the absolute income hypothesis developed by Keynes (1936). According to Keynes, consumption and saving are a functions of income. Income increases lead to higher consumption and savings. However, the marginal propensity to consumption and the marginal propensity to savings, which show the percentages of income to be allocated to consumption and savings, are determinative. While the marginal propensity to save is high in high-income holders, the marginal propensity to consume is high at the low-income levels. According to Keynes, with a rise in income, the percentage of income allocated to savings increases. Empirical studies initially proved Keynes' absolute income hypothesis by finding a positive relationship between the share of income saved and income. However, the study conducted by Kuznets for the USA in the period after 1899, reached findings that conflicted with Keynes' theory and it was determined that the share allocated to savings from income remained the same despite the increase in real income (Friedman 1957: 3-4; Modigliani 1986: 705). These empirical contradictions have led economists to develop new approaches to explain consumption and savings.

The first of the new approaches is Duesenberry (1967)'s relative income hypothesis. The hypothesis assumes that Keynes' absolute income hypothesis is incomplete since it assumes that each consumer's consumption expenditure decisions are independent of other consumers and the consumption relations are reversible. In this regards, the Keynesian theory serves as a special case of the general consumption theory (Duesenberry 1967: 1). The relative income approach depends on the explanation of Brady and Friedman's view that a consumer's consumption is independent of absolute income, however, in terms of imitation of upper-income groups (Modigliani 1986: 705) it is dependent on the position of the consumer in the distribution of income among consumers in the community (Friedman 1957: 4). According to Dusenberry's relative income hypothesis, utility is a function of relative income rather than absolute income (Dusenberry 1967: 112). The behavior of those at the upper limits of the income distribution differs in two ways compared to those at the lower limits. First, the change in saving rates affects total savings much less, secondly, the type of consumption in high-income groups is more affected by the competitive considerations than in lower-income groups (Duesenberry 1967: 113). In addition, Duesenberry's approach makes it possible to interpret aggregate data by expressing the ratio of consumption to income as a function of the ratio of current income to the highest level it reaches earlier (Friedman 1957: 4).

The inconsistency that emerged in Keynes' theory because consumption and savings show a certain regularity despite changes in current income was tried to be resolved by Modigliani and Brumberg's lifetime income hypothesis developed in 1954. (Modigliani 1986: 705). Modigliani and Brumberg (2005) obtained results that are basically compatible with Keynes' theory. The main difference is that instead of the psychological factor called animal motives in Keynes, they treat people as having forward-looking expectations (Modigliani and Brumberg 2005: 32). According to the lifetime income hypothesis, saving is proportional to the average income-earning capacity of households rather than their current income, as it is made to provide a cushion against large changes and short-term fluctuations in income over the life cycle (Modigliani and Brumberg 2005: 32). The ratio of income to savings is independent of current income, and the deviations result from short-term fluctuations in households' earning capacity and changes in this capacity (Modigliani and Brumberg 2005: 32). According to the lifetime income hypothesis since the income will be high in the working age, the saving will also be high, and dissaving will occur in the retirement period (Rijckeghem and Üçer 2009: 20). Basically, consumers are making an intertemporal smoothing.

Friedman developed the permanent income hypothesis which makes a distinction between recorded income and permanent income and explains consumption and saving behaviors according to the latter (Friedman 1957: 221). Temporary changes in a consumer's income would not have an impact on consumption unless they transform into permanent effects. In this case, while consumers are not sensitive to temporary shocks in their incomes, they smoothen by adjusting their consumption considering permanent effects. The permanent income hypothesis implies Ricardian equivalence, which suggests that private savings and public savings would balance. Increases in public savings with tax rises cause a decrease in private savings and reductions in public savings due to expenditure cause an increase in private savings (Rijckeghem and Üçer 2009: 21).

The studies in the empirical literature can be classified into two specific groups. The first group analyzes studies cross-countries. Edwards (1995) analyzes 32 countries and finds that the determinants of public and private savings vary. While demographic variables, social security expenditures and the depth of the financial sector do not affect public savings, private savings are sensitive to these variables. Opoku (2020) explores the income and substitution effects of short-term nominal interest rates in 19 OECD countries. The findings show that the substitution effect outweighs the income effect in the long run and the short run, the income effect is higher than the substitution effect. Inflation, wealth, income and wealth taxes, unemployment rate and general government gross debt have negative effects on household savings in the long run. Hunt et al. (2021) find that in 36 OECD countries, in addition to traditional determinants of household savings such as life expectancy and income tax rate, changes in socio-economic and demographic conditions are also influential. A narrower gender gap in access to higher education and employment leads to a larger household savings rate. Fredriksson and Staal (2021) examine 14 OECD countries and identify the positive effects of unexpected income changes and unexpected inflation on savings. Uncertainty affects savings positively, while social security suppresses savings.

The second group studies focus on a certain country. While these studies deal with demographic variables, they also examine other effects such as cultural factors and macroeconomic variables. Finlay and Price (2015) find that saving behavior varies between age groups in Australia. They find that situations that increase risk, such as being a single-parent and migrant household, are negatively associated with savings. They also find that savings are positively correlated with incomes and negatively correlated with wealth and gearing. Mirach and Hailu (2014) find that demographic factors such as age, gender, marital status, as well as the existence of financial institutions where savings will be used, and cultural background also affect savings in Ethiopia. Rehman et al. (2011) investigate the determinants of savings in Pakistan for different income groups. While saving increases with income in low and middle-income groups, children's education expenditure, family size, and household obligations per capita are negatively related to saving. In the higher income group, the findings are consistent with the lifetime income hypothesis. Pan (2016) examines the savings of rural and urban households in China. Savings in rural areas are largely explained by income. In addition, having a school-age child is also decisive in high-income quantiles. Changes in quantile regression coefficients explain the urban saving rates. Paiva and Jahan (2003) find that private savings and public savings are offsets in Brazil. Private savings have a high and inverse response to public savings. In addition, financial depth and terms of trade positively affect savings. Curtis et al. (2015) examine the effect of demographic changes on household savings in China. Demographic changes explain more than half of household savings rates.

Among the studies on Türkiye, Ozcan et al. (2003) find that a change in one of the determinants of saving is effective in the long run rather than the short run. Public savings crowd out private savings. The increase in public savings is balanced by the decrease in private savings. The income level has a positive effect on the private saving rate. The negative impact of life expectancy supports the life cycle hypothesis. Terms of trade shocks increase private savings. The economic crisis affects the savings rate negatively. Nalın (2013) finds that inflation can increase household savings in Türkiye if other macroeconomic factors remain constant. Household income, education level, occupation, place of residence (rural/urban), car ownership, and household size are other important variables in explaining the change in household savings and portfolio preference behavior. Rijckeghem (2010) examines the decline in savings in Türkiye and finds being a homeowner is decisive. Homeowners substantially reduced their savings rates, while the decrease in tenants was minor. Households with interest income do not reduce their savings.

This study examines the effects of household characteristics on household savings in Türkiye. Saving rates are crucial determinants of many variables and economic development, and the examination of savings rates is extremely important for contributing to the literature and policy implications.

Figure 1 presents household savings rates which fluctuate at a low level in Türkiye. Savings ratios, which fell below 10% in the post-global crisis period, rose between 2013-2017 and declined in the last four years.

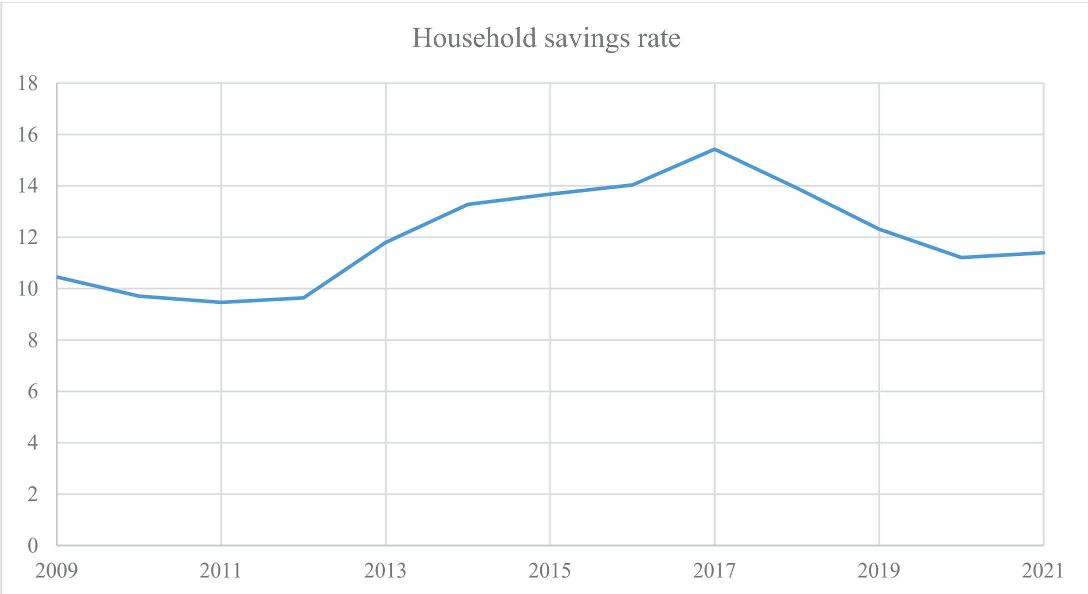

Figure 1. Household Savings Rates in Türkiye

Source: Compiled from Turkish Statistical Institute report (TURKSTAT 19.02.2023)

The most important possible explanation for fluctuations in savings rates is income level. It is theoretically consistent that savings fell until 2012 due to the economy that shrank by 5% in the global crisis. It is possible to explain the increase in savings with the recovery of the economy and the start of growth in the 2013-2017 period. The recent decline in savings may have contributed to the increase in the inflation rate. It is theoretically and empirically known that households increase their expenditures to protect themselves from price increases in the future in high inflation periods.

The next section explains the dataset and the method. Then the findings are presented and interpreted. The last section summarizes the results and discusses policy recommendations.

2. DATA AND METHODOLOGY

In this study, the logistic quantile regression method was employed to determine the household characteristics that affect household savings rates in Türkiye. The data set of the study was obtained from the 2019 Household Budget Survey applied by the Turkish Statistical Institute (TURKSTAT). Data for 2019 were preferred to avoid the impact of the pandemic on the household budget balance. Household Budget Survey was applied to 15552 households across Türkiye in 2019, and responses were received from 11521 households. In this study, the factors affecting savings rates, especially for homeowners and tenants, were examined and compared. Therefore, the sample size consists of 9669 households, of which 7092 homeowners and 2577 tenants. The budget survey is applied separately for households and individuals. Here, the data of the questionnaire applied to the households were used and the characteristics of the households were examined. The dependent variable, the annual savings rate of the household was calculated as the ratio of annual savings to annual disposable income. Among the independent variables, the logarithmic annual disposable income was the main variable focused on, and the change in the effect of this variable on the saving rate at the level of quantiles was graphically examined. In addition, other characteristics of the house and the household were employed in the model, and the model findings that had a statistically significant effect on the savings rate were interpreted. The explanations and details of the variables included in the analysis are given in Table 1.

Table 1. Variables and Definitions

Savings_Rate	Annual savings / annual disposable income (Dependent)
Log_Inc	Logarithmic annual disposable income
No_Hh	Total number of members living in the household
Size_H	Size of the residence (10 m2)
Calorifere	Calorifere ownership
Sec_H	Second home ownership
No_Mob	Number of mobile phones
No_Pc	Number of computers
No_Net	Number of internets
No_Oto	Number of cars (excluding those for commercial use)
Smoking	Presence of individuals in the household who have the habit of smoking cigarettes, tobacco and cigars
Alcohol	Presence of individuals in the household who have the habit of using alcoholic beverages
Eat_Out	Presence of the habit of eating lunch or dinner out
Paid_Spor	Presence of individuals engaged in sports, entertainment, culture, etc. activities by paying a fee in the household
Paid_Tv	Presence of paid TV subscriptions in the household
Cafe	Presence of individuals in the household who have the habit of going to coffee houses, cafés, etc.
Ccard	Presence of individuals using credit cards in the household
Market	Presence of the habit of going to the market in the household
Onl_Shop	Presence and frequency of online shopping habits in the household

Dummy variables (Calorifere, Sec_H, Smoking, Alcohol, Eat_Out, Paid_Spor, Paid_Tv, Café, Ccard, Market, and Onl_Shop) are coded so that a value of zero represents absence. Summary statistics for household savings rate and logarithmic income for homeowners and tenants are given in Table 2. When the table is explored, it could be seen that nearly 75% of the households in the country-wide sample are homeowners. When the income level is examined, it is determined that the average income level is close for the homeowners and the tenants.

Table 2. Summary Statistics for Household Savings Rate and Income

	Variables	Number of Observations	Mean	Median	Min	Max
Home owner	Savings_Rate	7092	0.0452	0.1599	-12.506	0.9567
	Log_Inc	7092	10.882	10.876	7.9399	13.791
Tenant	Savings_Rate	2577	-0.0733	0.0722	-35.806	0.8475
	Log_Inc	2577	10.791	10.796	6.2328	14.219

It can be suggested that the lowest and highest income levels are similar for both groups. On the other hand, the two groups present a significant distinction in saving rates. The average savings rate for homeowners is positive and around 4.5%, while the average savings rate for tenants is negative. In other words, it is seen that the tenants are in debt on average. The fact that the median values are significantly larger than the mean indicates that the savings rates have an asymmetrical distribution and are skewed to the left in both groups. This inference, which was obtained by the location measures of the distribution of savings rates for both groups, was supported by both graphical methods such as histograms and normal distribution tests. As a result of the examinations and tests, the asymmetrical distribution of savings rates in both groups was determined clearly. The presence of large negative values in minimum savings rates indicates the presence of households with high debt levels and extreme value characteristics. However, the maximum savings rate is limited to 1 from the upper. The saving rate variable, which shows the continuous, limited, and asymmetric distribution, has a suitable structure for analysis with logistic quantile regression.

In this study, the logistic quantile regression method is utilized to investigate household savings rates. The quantile regression is developed as an alternative method to the classical mean-based regression models since it is based on modeling the conditional quantiles of the dependent variable rather than the conditional mean. The quantile regression method is more robust to the existence of outliers and allows for investigation of the targeted point of the dependent variable distribution by allowing to stretch of the normality assumption in classical models. The logistic quantile regression model was presented by Bottai et al. (2010) as an alternative approach to the quantile regression model. This approach is proposed for models with a limited and continuous dependent variable within a certain range. The quantile of a variable is invariant under monotonous transformations. In other words, for a non-decreasing function , the q. quantile of the variable y, $Q_y(q)$, has the feature of

$$h\left(Q_y(q)\right) = Q_{h(y)}(q) \tag{1}$$

However, the expected value of the variable, and therefore its mean, does not have this feature. Based on this feature of the quantile regression, it has been proposed to estimate the conditional quantiles after applying a monotone transformation to the variable if the dependent variable is limited. This transformation, called the link function, is preferred as the logistic transformation in logistic quantile regression:

$$h(y_i) = \log\left(\frac{y_i - y_{\min}}{y_{\max} - y_i}\right) \tag{2}$$

Here, the $(y_{\min}, y_{\max})$ values do not have to be the lowest and highest values of the variable but are the values that limit the variable from above and below. This transformation, which is applied to the probability values in the range $(0,1)$ in the logistic regression, is applied here for the continuous and limited dependent variable. The quantile regression model is estimated using the transformed dependent variable:

$$Q_{h(y_i)}(q) = Q_{\mathrm{logit}(y_i)}(q) = x'_i \beta_q \tag{3}$$

Here, β_q represents the regression coefficients for the $q.$ quantile of the dependent variable. For the coefficients, bootstrap standard errors, which are more successful, are used instead of asymptotic standard errors (Bottai et al. 2010).

3. FINDINGS

Before examining the determinants of household saving rates at the quantile levels, assumptions were tested in the classical mean-based regression model. The OLS model results estimated separately for both homeowners and tenants will be presented along with the logistic quantile regression results. However, before moving on to the estimation results, heteroscedasticity and normality tests were applied for the residuals of the OLS models. The results of the Breusch-Pagan Test applied for heteroscedasticity and the Jarque-Bera Test applied for normality are given in Table 3.

Table 3. Heteroscedasticity and Normality Tests Results for OLS Models

	Test	Test Stat. (Chi2)	Prob.
Home owner	Breusch-Pagan	1426.43	0.000
	Jarque-Bera	2.8e+08	0.000
Tenant	Breusch-Pagan	41354.63	0.000
	Jarque-Bera	1.5e+08	0.000

When the Breusch-Pagan Test test results are examined, the null hypothesis of the test suggesting that the homoscedasticity is valid, is rejected according to the tail probabilities and there is a problem of heteroscedasticity in the OLS models. Similarly, as a result of the Jarque-Bera Test with the null hypothesis that the residuals are normally distributed, the null hypothesis was rejected in both groups according to the tail probabilities, and the OLS residuals do not have a normal distribution. Therefore, the assumptions of the OLS model are not provided and it is appropriate to prefer the quantile regression.

In the next step, the savings rates for both homeowners and tenants are modeled in low, medium and high quantiles. For this purpose, the models were estimated at 25th, 50th (median) and 75th quantiles, respectively. Table 4 presents OLS and logistic quantile model estimation results. The results suggest that the main determinant of household savings rates for both homeowners and tenants is household income. The income variable is considered as the main determinant of expenditures and saving rates according to the economic theory. The results of the model report that coefficients of the income variable in all models are significantly higher compared to the coefficients of all other variables. As a result, inline with the economic theory, it has been revealed that income is the most influential variable on the change in saving rates among the variables examined in the model.

Table 4. Model Estimation Results

	Homeowners				Tenants			
Variables	**OLS**	**q25**	**q50**	**q75**	**OLS**	**q25**	**q50**	**q75**
Log_Inc	0.505***	0.515***	0.544***	0.609***	0.905***	0.608***	0.573***	0.582***
Size_H	-0.007***	-0.010***	-0.009***	-0.008***	-0.025***	-0.015***	-0.014***	-0.007**
Calorifere	-0.090***	-0.080***	-0.113***	-0.131***	-0.233***	-0.111***	-0.122***	-0.116***
No_Mob	-0.017**	-0.019*	-0.028**	-0.038***	-0.067***	-0.024	-0.020	-0.028**
No_Pc	-0.032***	-0.020	-0.029**	-0.045***		-0.013	-0.008	-0.036***
No_Oto	-0.179***	-0.201***	-0.141***	-0.111***	-0.334***	-0.230***	0.155***	-0.159***
Smoking	-0.046***	-0.026*	-0.055***	-0.083***		-0.078***	-0.061***	-0.090***
Alcohol	-0.101***	-0.147***	-0.151***	-0.162***	-0.114**	-0.042	-0.062*	-0.097***
Eat_Out	-0.048***	-0.051***	-0.064***	-0.083***	-0.141***	-0.103***	-0.110***	-0.101***
Paid_Spor	-0.101***	-0.095**	-0.133***	-0.115***	-0.158***	-0.111***	-0.132***	-0.146***
Paid_Tv	-0.085***	-0.086***	-0.068***	-0.117***	-0.162***	-0.139***	-0.108***	-0.112***
Cafe	-0.049***	-0.056***	-0.036***	-0.029*	-0.116***	-0.115***	-0.074***	-0.076***
Ccard	-0.057***	-0.046**	-0.060***	-0.100***	-0.107***	-0.098***	-0.090***	-0.065***
Onl_Shop	-0.040***	-0.057***	-0.048***	-0.067***	-0.114***	-0.084***	-0.056***	-0.053***
No_Hh	-0.021***	-0.013**	-0.021***	-0.024***	-0.043***	-0.035***	-0.042***	-0.042***
Sec_H	-0.068***	-0.066**	-0.059***	-0.058***				
No_Net	-0.026***	-0.026***	-0.023***	-0.019***				
Market	-0.070***	-0.082***	-0.069***	-0.107***	-0.065*			
Cons	-4.898***	-1.496***	-1.495***	-1.840***	-8.677***	-2.348***	-1.761***	-1.644***

Note: *, ** and *** indicate statistical significance for 10%, 5% and 1% margin of error, respectively.

It has been determined that the impact of income on the savings rate is generally higher for tenants than for homeowners. However, as a finding that could not be obtained with the OLS model, it was determined that the effect of income at the quantile levels was non-linear and showed a quadratic trend. The change in the effect of income at the levels of quantiles is interpreted in the graph in Figure 2.

When the effect of other variables is examined, it is seen that the variables, which are generally expenditure items, have a significant effect on all quantiles and negatively affect the savings rate. While the effect of the size of the house and the number of cars, which are variables that have a significant effect on both the homeowners and the tenants, decreases towards the higher quantiles, the effect of the number of people living in the household increases. Credit card ownership has different effects on homeowners and tenants. In the case of the move from low savings to higher, the negative impact of credit card ownership increases for homeowners and decreases for tenants. The negative effects of calorifere ownership, habits of smoking, alcohol, and eating out on the savings rate increase linearly across the quantiles for homeowners. Cafe habit has a linearly decreasing effect along the quantiles for homeowners. However, the effect of the same variables differs for the tenants in the extreme quantiles relative to the median and is not linear across the quantiles. While alcohol habit does not have a significant effect on the savings rate at low quantiles for tenants, it becomes more significant towards higher quantiles and its negative effect on the savings rate becomes stronger. The effects of paid TV and paid sports habits are also non-linear for homeowners and differ in extreme quantiles. There are variables with different effects on homeowners and tenants. The number of mobile phones in the household has a significant effect on all quantiles for homeowners, and its effect becomes stronger as move towards higher savings rates. For tenants, it has a significant effect only on households with high savings rates. While the number of computers in the household does not have a significant effect on savings rates for both homeowners and tenants in low quantiles, it has a negative and significant effect in high quantiles. In addition, it has a significant effect on the median, that is, the medium savings level, for homeowners, but not for tenants. On the other hand, while the variables of the number of computers and smoking habits were not found significant for tenants in the OLS model, they were found to have a significant effect when analyzed at the quantile levels. The opposite is true for market habits.

In addition, some variables have an impact on savings rates for homeowners but not for tenants. These variables are second home ownership, the number of internet and market habit. These variables only have a negative effect on savings rates for homeowners. Second home ownership and the number of internet variables were excluded from the tenants model because, unlike homeowners, they had no statistically significant effect on either the average or the quantiles of savings rates. This is because although these variables are key expenditure items for homeowners, most tenants do not have these. The Markets variable, on the other hand, was excluded from the quantile models, although it had an effect on the average savings rate for tenants, as it did not have a statistically significant effect at the quantiles level. It has been determined that although the market habit has an effect on the average savings rates for tenants, it is not a determining factor in low and high savings rates.

It has been determined that the effect of household income on the saving rate is not linear across the quantiles. The change in the income effect across quantiles for homeowners and tenants is shown in Figure 2.

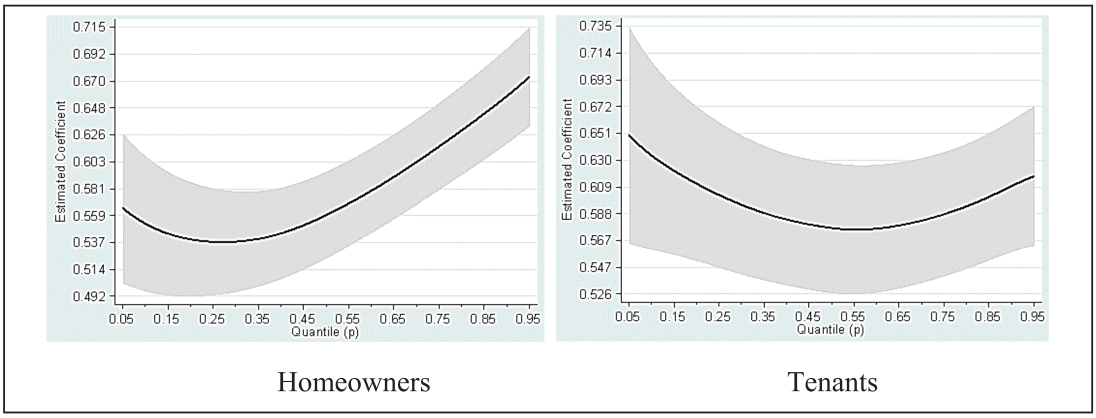

Figure 2. Change of the Effect of Household Income across the Quantiles

Examining the graphs, it is understood that income has a quadratic effect across quantiles for both homeowners and tenants. However, the change in the impact differs for homeowners and tenants. It is seen that the positive effect of income decreases for homeowners until the 30th quantile, in other words, for the households in the lowest 30 percent according to the savings rate. However, from this point on, as the savings rate increases, the positive effect of income gets stronger. It has been determined that this turning point for the tenants is around the 60th quantile. In other words, while the positive effect of income weakens as the savings rate rises for households in the low 60 percent, it increases for the 40 percent with high savings rates. However, this increase is not as sharp as for homeowners. Income is most effective in households with the highest savings rate for homeowners, and the lowest savings rate for tenants. The lowest income effect generally occurs in households with average savings.

4. CONCLUSION

The factors that determine household savings rates are the subject of extensive research in the literature. In particular, the relationship between income level, which is the main determinant of saving, and saving is important. This study analyzes the determinants of household savings rates and discusses the relationship between income and savings in detail. OLS model estimations lose their reliability due to the presence of extreme values in the dependent variable. The quantile regression models allow us to examine the desired point of the distribution instead of the conditional mean of the variable being studied. Logistic quantile regression models, on the other hand, have been proposed as an alternative to quantile models when the dependent variable is continuous and limited within a certain range. In this study, unlike previous studies in the literature, savings rates were examined with logistic quantile regression models. Since the household savings rate is limited to 1 from the upper, it is suitable for these models. In addition, separate models were estimated for homeowners and tenants, and differences at the level of quantiles were revealed.

When the findings were examined, it was determined that the principal determinant of the savings rate was income level for both homeowners and tenants. Income level has a positive effect on the saving rate. However, this effect is not linear across the quantiles and differs between homeowners and tenants. When the change of the income effect according to the quantiles is examined, it is seen that there is a quadratic change throughout the quantiles. Accordingly, the effect of income decreases up to a certain saving rate and then increases. However, this turning point occurs quicker for homeowners than for tenants, with a stronger impact afterward. Income is more effective at high savings rates for homeowners, while it is more effective at low savings rates for tenants.

When the other determinants of the savings rate are examined, it has been determined that the variables are generally expenditure items and have a negative effect on the savings rate. It is understood that the main expenditure items affect the savings rate for both homeowners and tenants. However, while some of these effects are linear throughout the quantiles, most are non-linear. When the coefficients of the quantile models are examined, it is seen that the effects of these variables do not increase or decrease linearly from low quantile to high. This is because the variables examined have different effects on low and high savings rates. Because the priority order and amount of income and expenditure items differ for families with different savings rates. When the order of priority or amount of these items changes, their impact on savings rates may decrease up to a threshold and then start to rise and vice versa. Therefore, there is a nonlinear change at the quantiles level. In this case, examining the savings rates at the level of quantiles rather than the average allows to obtain more detailed and realistic findings. It has been understood that expenditure items are generally more effective in extreme quantiles, that is, in low and high savings rates than the average.

High savings rates are critical for many economic variables. With the attainment of high savings rates, the capital stock and growth rates will increase in the Turkish economy. High savings will also allow the current account deficit to decrease. Therefore, policy recommendations are crucial to increase savings. The findings show that the most significant determinant for the high saving rate is income level. Although the Turkish economy achieves high growth rates from time to time, these periods are interrupted by domestic or foreign crises. For this reason, the most crucial policy proposal is to ensure long-term growth that is not interrupted by crises.

The negative impact of all expenditure items on savings shows the importance of inflation on savings rates. High inflation reduces the savings rates by pushing the expenditures forward and increasing the share of expenditure items in the budget. So the second policy recommendation is to have low and stable inflation.

Funding

This research did not receive any specific grant from the public, commercial, or not-for-profit funding agencies.

Conflict of Interest

All authors declare that there is no conflict of interest.

Submission Declaration Statement

We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere.

REFERENCES

- BOTTAI, M., CAI, B., and MCKEOWN, R.E. (2010). Logistic Quantile Regression for Bounded Outcomes. *Statistics in Medicine*, 29(2), 309-317.

- CURTIS, C.C., LUGAUER, S., and MARK, N.C. (2015). Demographic Patterns and Household Saving in China. *American Economic Journal: Macroeconomics,* 7(2), 58–94.

- DOMAR, E. (1946). Capital Expansion, Rate of Growth, and Employment. *Econometrica,* 14(2), 137–147.

- DUESENBERRY, J.S. (1967). *Income, Saving, and the Theory of Consumer Behavior.* New York: Oxford University Press.

- EDWARDS, S. (1995). Why Are Saving Rates So Different Across Countries?: An International Comparative Analysis. *NBER Working Paper Series.* No. 5097.

- FINLAY, R. And PRICE, F. (2015). Household Saving in Australia. *The B.E. Journal of Macroeconomics,* 15(2), 677–704.

- FREDRIKSSON, C. and STAAL, K. (2021). Determinants of Household Savings: A Cross-Country Analysis. *International Advances in Economic Research,* 27, 257–272.

- FRIEDMAN, M. (1957). *A Theory of the Consumption Function.* Princeton University Press, ISBN: 0-691-04182-2.

- HARROD, R. F. (1939). An Essay in Dynamic Theory. *The Economic Journal, 49(193), 14–33.*

- HUNT, E., JEON, H., and LEE, S. (2021). Determinants of Household Savings: An Empirical Evidence from the OECD Member Countries. *Business and Economic Research,* 11(2), 62-75.

- KEYNES, J. M. (1936*), The General Theory of Employment, Interest, and Money.* London: MacMillan. ISBN-10: 0333107292

- LUCAS, R.E. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics.* 22, 3–42.

- MIRACH, T.H. and HAILU, Y.M. (2014). Determinants of Household Saving in Ethiopia- The Case of North Gondar Zone, Amhara Regıonal State. *International Journal of Development and Economic Sustainability,* 2(4), 37-47.

- MODIGLIANI, F. (1986). Life Cycle, Individual Thrift, and the Wealth of Nations. *Science*. 234(4777), 704-712.

- MODIGLIANI, F. and BRUMBERG, R. (2005) Utility Analysis and The Consumption Function: An Interpretation of Cross-Section Data. in *The Collected Papers of Franco Modigliani.* Ed: F. Modigliani. 3-45, ISBN: 0-262-13150-1.

- NALIN, H.T. (2013). Determinants of Household Saving and Portfolio Choice Behaviour in Turkey. *Acta Oeconomica,* 63(3), 309–331.

- OPOKU, P.K. (2020). The Short-Run and Long-Run Determinants of Household Saving: Evidence from OECD Economies. *Comparative Economic Studies,* 62, 430–464.

- OZCAN, K.M., GUNAY, S., and ERTAC, S. (2003). Determinants of Private Savings Behaviour in Turkey. *Applied Economics,* 35, 1405–1416.

- PAIVA, C. and JAHAN, S. (2003). An Empirical Study of Private Saving in Brazil. *Brazilian Journal of Political Economy.* 23(1), 124-137.

- PAN, Y. (2016). Understanding the Rural and Urban Household Saving Rise in China. *Regional Science and Urban Economics,* 56, 46-59.

- REHMAN, H. ur, BASHIR, F., and FARIDI, M.Z. (2011). Saving Behavior among Different Income Groups in Pakistan: A Micro Study. *International Journal of Humanities and Social Science,* 1(10), 268-277.

- RIJCKEGHEM, C.V. (2010). Determinants of Private Saving in Turkey: An Update. *Boğaziçi Üniversitesi İktisat Bölümü Çalışma Tebliği,* No.4.

- RIJCKEGHEM, C.A. and ÜÇER, M. (2009). The Evolution and Determinants of the Turkish Private Saving Rate: What Lessons for Policy? *Koç University-TÜSİAD Economic Research Forum.*

- SOLOW, R.M. (1956). A Contribution to the Theory of Economic Growth. *Quarterly Journal of Economics,* 70(1), 65–94.

- TURKSTAT. (2022). Macro Economic Indicators, Current Prices, 2009-2021 [online]. *Institutional Sector Accounts*, https://data.tuik.gov.tr/Bulten/Index?p=Kurumsal-Sektor-Hesaplari-2021-45684, [Date Accessed: 19/02/2023].

JAME
E-ISSN:2791-7401
Volume : 3, Issue 1, 2023
URL: https://journals.gen.tr/jame

Received: 2023-04-25

Accepted: 2023-06-15

RESEARCH ARTICLE

Comparison of GARCH and SVR-GARCH models: Example of gold return

Zeynep ŞENGÜL

Marmara University, Faculty of Economics, Department of Econometrics, PhD Student, Türkiye, e-mail: zzeynepsengul@gmail.com

Abstract

Gold has been a precious resource for people on earth from the past to the present. It is used as both a value gain and jewelry, and is the focus of interest for people in terms of receiving attention and protecting its value. Especially recently, it has been the most favorite for investors due to its excess value increase and decrease which is constantly monitored. The study aimed to compare the predictive performance of the gold price return using the Support Vector Regression-GARCH hybrid models combined with the traditional volatility models. It has been examined whether the Support Vector Regression GARCH models would increase foresight performance. The study used data on the daily frequency between 01/01/2010–01/04/2023. Generalized Autoregressive Conditional Variable Variance, Glosten-Jaganthan-Runkle GARCH, Exponential GARCH and hybrid model Support Vector Regression -GARCH are utilized as prediction methods. For all methods, the gold series is divided into two groups as training and test data. The Root Mean Square Error values are compared as a model performance criterion. The RMSE values and graphics outputs have been concluded that the Support Vector Regression-GARCH model based on predicted linear, radial-based and polynomial kernel predicts more effectively than the GARCH models.

Keywords: Financial Market, GARCH Models, SVR-GARCH Model, Machine Learning, Time Series

JEL Codes: C22, C53, C58

Citation: ŞENGÜL, Z. (2023). Comparison of GARCH and SVR-GARCH models: Example of gold return. *Journal of Applied Microeconometrics (JAME)*. 3(1), 23-35, DOI: 10.53753/jame.2068

Corresponding Author:
Zeynep ŞENGÜL
E-mail: zzeynepsengul@gmail.com

1. INTRODUCTION

Since gold has been considered as an investment tool from the past to the present, it attracts the attention of individuals and institutions due to its exchangeability. Especially in recent times, the price value of gold has encouraged investors to follow gold more. Gold is constantly volatile likewise other financial instruments. This is due to the issues that occur within the country as well as in the globe. Therefore, it is important to model this volatility that financial markets have. This is because investors would like to make maximum profit from their investments. However, the high volatility of financial markets means that the risk is high. As a result of the high risk, investors want to have an insight into the financial assets they would invest in before making the actual investment.

The traditional methods (ARIMA, GARCH, etc.) are used to estimate volatility in their financial series. When the literature is examined, it is stated that recently machine learning algorithms are included in the predictions of the financial time series and making effective forecasts (Bildirici & Ersin 2012).

The econometric model commonly used for estimating volatility in financial markets is the ARCH model developed by Engle (1982). However, the ARCH model assumes that the impact of positive and negative news affecting the market on volatility is the same. Due to the diversity of the data and the differing problems, the ARCH model is developed by Bollerslev (1986) and is called the GARCH model. The ARCH and GARCH models assume that the variance effect of positive and negative shocks is the same. However, it seems that negative shocks representing bad news in the financial markets affect volatility more than positive shocks which represent good news. For this reason, the E-GARCH model, expressed as the exponential GARCH (E-GARCH), has been developed by Nelson (1991) to eliminate weaknesses ignored in the symmetrical models (Engle 1993: 75; Nelson 1991). One of the main shortcomings of the GARCH model is that this model does not consider the possible asymmetrical impact observed in the financial time series. Therefore, the GJR-GARCH model has been proposed by Glosten, Jaganathan and Runkle (1993), which takes into account the asymmetrical impact.

The study will estimate the GARCH, GJR-GARCH, E-GARCH and SVR-GARCH models and use the RMSE value determined as the model performance criterion. This study aims to examine whether the SVR-GARCH hybrid model is associated with higher performance compared to the GARCH, GJR-GARCH and E-GARCH models.

2. LITERATURE

When the literature is examined, Perez-Cruz et al. (2003) stated that the models predicted using SVR, in contrast to the GARCH models performed with the most probability estimation method applied to the time series in their studies, performed the best prediction.

Alberg et al. (2008) used GJR-GARCH, APARCH, E-GARCH models to estimate the return and conditional variance on the Tel Aviv Stock Exchange. The study utilized a variety of comparison criteria (i.e. MSE, MedSE, MAE and AMAPE and TIC) for comparison purposes and determined that the E-GARCH model with student-t distribution was the best predictor.

Ou and Wang (2010) used LS-SVM (Least Square Support Vector Machine), GARCH, E-GARCH, GJR-GARCH models to estimate the volatility of the ASEAN stock exchange. They stated that the LSSVM model provided more resistant and robust performance against volatility.

Jena and Goyari (2010) investigated the existence of a high volatility regime between 2005 and 2009, using the MS-ARCH model for oil and gold prices traded in the Indian market. The study reported that the high volatility was observed during the global financial crisis and the crisis is passed to a lower volatility regime after the crisis.

Bildirici and Ersin (2012) estimated BIST-100 index using the GARCH, SVR-GARCH and MLP (artificial neural networks)-GARCH models. It was concluded that SVR-GARCH and MLP-GARCH were better than the GARCH model.

Wang et al. (2013) compared error statistics for the model performance benchmark by performing Markov-switching (MSM), GARCH (1,1) and SVM-based Markov-switching (SVM-MSM) model forecasts for two different financial time series. The results showed that the best models were SVM-MSM, MSM and GARCH (1,1) respectively.

Gürsoy and Balaban (2014), using the BIST-100 index, estimated GARCH, E-GARCH, GJR-GARCH and SVR-GARCH models and stated that the best model was the SVR-GARCH model.

Karabacak et al. (2014) predicted volatility with the ARCH, GARCH, TARCH, E-GARCH and IGARCH models using the BIST-100 index return and gold return series in their study. TARCH stated that the best model for BIST 100 index return was the GARCH model for the gold return series.

Birau et al. (2015) used Bombay Stock Exchange Bank Index (BANK) used ARCH and GARCH models for volatility estimation. It was stated that the GARCH model predicted better than the ARCH model.

Katsiampa (2017) employed GARCH, EGARCH, TGARCH, APGARCH, C-GARCH, and AC-GARCH models for volatility modeling using Bitcoin data in his study. By comparing the AIC, SIC and HQ information criteria of the models, he determined that C-GARCH was the best model.

Cihangir and Uğurlu (2017) used GARCH, APARCH, TARCH and EGARCH estimation methods for the volatility of the gold price between 2010 and 2016 for their work. The study reported that APARCH was the most appropriate model**.**

Peng et al. (2018) examined the volatility of three different cryptocurrencies they identified. They applied GARCH, EGARCH, GJR-GARCH and SVR-GARCH models to the daily and hourly frequency data. The study reports that the SVR-GARCH model performs better than other models.

When the literature is examined econometrically, it is seen that gold prices and GARCH models are included in many studies. In this study, the hybrid model created by using the traditional methods and integrating with SVR, the recently widespread machine learning prediction algorithm, is created. For the performance evaluation of the applied prediction models, it is aimed to select the best prediction model by comparing the Root Mean Square Error (RMSE) values.

3. METHODS OF RESEARCH

Modeling and predicting volatility is very important in the financial markets. The high volatility states that the financial asset is risky. The correct estimate of the financial return volatility is crucial to assess investment risk.

Linear and nonlinear methods are used in time series. Linear models could be listed as ARIMA and GARCH. Support Vector Regressions and Artificial Neural Networks can be exemplified to nonlinear models. In this study, the hybrid GARCH model combined with SVR with the GARCH, GJR-GARCH, E-GARCH models will be estimated and the model with small error statistics will be determined.

The ARCH model is an autoregressive model developed by Engle (1982). ARCH is one of the most widely used models to model volatility in the market. It assumes that the variance value of the error terms is related to the previous period error values. It is possible to predict the variance of the series within a given time.

The ARCH model does not allow the conditional variance to change over time as a function of past errors which leaves the unconditional variance constant (Bolleslev 1986).

The ARCH model formulas are showed in Equation 1:

$$
\begin{aligned}
&y_t = \theta_0 + \theta_1 y_{t-1} + \cdots + \theta_n y_{n-1} + \varepsilon_t \qquad\qquad \varepsilon | I_{t-1} N(0, \sigma_t^2) \\
&\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 \\
&\varepsilon_t = y_t - x_t b
\end{aligned} \tag{1}
$$

ARCH model constraints exist; h_t and ε_t must be positive. In other words, $\alpha_0 > 0$ other parameters $\alpha_i \geq 0$. The other constraint should be $0 \leq \alpha_i \leq 1$ (Engel 1982).

The GARCH model is developed by Bolleslev (1986) and Taylor (1986). The GARCH model estimates the conditional variance of the process variable based on its own delayed values. The error squares calculated in the mean equation gives information about volatility in past periods. When the literature is examined, it is seen that the GARCH model makes a more effective prediction than the ARCH model. The GARCH (1,1) model is the simplest but most powerful model of volatility (Engle 2001).

The GARCH model formula is showed in Equation 2:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \tag{2}$$

As a GARCH model constraint is, $\alpha_0 > 0,\ \alpha_i \geq 0\ ,\ \beta_i \geq 0$ and $\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$ (Bolleslev 1986).

The GJR-GARCH model is developed in 1993 by the Glosten, Jaganthan and Runkle. This model reacts to past negative and positive changes of the conditional variance. This model is recommended to determine the leverage effect in time series.

The GJR-GARCH model formula is showed in Equation 3:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \gamma_i \varepsilon_{t-i}^2 I_{t-1} \sum_{i=1}^{p} \beta_j \sigma_{t-j}^2 \tag{3}$$

The I_{t-1} parameter in Equation 3 refers to unexpected news;

$$I_{t-1} = \begin{cases} 1 \text{ if } \varepsilon_{t-1} < 0, \text{ bad news} \\ 0 \text{ if } \varepsilon_{t-1} \geq 0, \text{ good news} \end{cases},$$

is expressed.

$\gamma \neq 0$ states that it has an asymmetric effect. The case where $\gamma > 0$ indicates that the leverage effect is present. Leverage means bad news is more effective than good news (Engle & Sokalska 2012).

The constraints of the GJR-GARCH model are $\alpha_0 > 0,\ \alpha_i > 0, \beta \geq 0, \gamma_i < 0$ and $\alpha_i + \gamma_i \geq$ and (Wang and Wu 2012).

The E-GARCH model is proposed by Nelson (1991), who added the leverage effect to the model to enable the asymmetric effect to be seen. The E-GARCH formula is showed in Equation 4:

$$log(\sigma_t^2) = \alpha_0 + \sum_{j=1}^{q} \beta_j \, log(\sigma_{t-j}^2) + \sum_{i=1}^{p} \alpha_i \left| \frac{\varepsilon_{t=i}}{\sigma_{t=i}} \right| + \sum_{k=1}^{r} \gamma_k \frac{\varepsilon_{t-k}}{\sigma_{t-k}} \tag{4}$$

The E-GARCH model responds asymmetrically to shock. Conditional variance is never negative due to logarithmic transformation in the model, it is always positive. The presence of $\gamma_k < 0$ in the model shows that the leverage effect exists (Çağlayan & Dayioğlu 2009). The advantage of the E-GARCH model allows unrestricted estimation of variance (Thomas & Mitchell 2005: 16).

SVR-GARCH is a powerful predictive method for predicting volatility and model risk. This is because the h_t output from the GARCH model is used as input in the SVR. The kernel determined in SVR provides effectiveness according to the structure of the data.

Instead of replacing the maximum likelihood method with SVR to predict GARCH parameters, it is recommended to combine the SVR and GARCH models to predict volatility. First, the GARCH model is used to obtain h_t. Then, $Z_t = f(Z_{t-1}, h'_{t-1}, y^2_{t-1}, h_{t-1})$, the nonlinear estimation is performed using the considering SVR model. In the equation is $Z_t = h'_t - h_t$. The linear GARCH model and the non-linear SVR model are combined to obtain estimates (Sun & Yu 2020).

In the SVR-GARCH model used to estimate volatility, the input vector is $x_t = [a^2, ht{-}1]$, and the output variable is *ht*. The SVR-GARCH structure is located below:

$$r_t = f(r_t{-}1) + a_t \tag{5}$$

Where f is the decision function predicted by SVR for the mean equation. After the squared residues from the conditional mean estimate of SVR-GARCH, estimate the conditional variance equation given below:

$$\tilde{h_t} = g(\tilde{h}_{t-1}, a^2_{t-1}) \tag{6}$$

In Equation 6, g is the decision function predicted by SVR. a^2_t refers to residual square and $\tilde{h}$ is the volatility. (Bezerra & Albuquerque 2017).

In the mean equation, we will use 3 different kernels. The kernel functions are included in Table 1.

Table 1. Kernel Functions

Kernel Function	Formulas
Linear	$K(x_i, x) = x_i^T x$
RBF	$K(x_i, x) = \exp(-\frac{1}{2\sigma^2} \|\| x - x_i^T \|\|^2)$
Polynominal	$K(x_i, x) = (x_i^T x + 1)^d \qquad d = 1,2 \dots, \infty$

The training set for determining the models and the test set are also used to evaluate the predictive performance of the models (Tay & Cao 2001). RMSE is used to evaluate predictive performance.

The RMSE value used to evaluate the effectiveness of the models and provides information concerning the active predictor with the least error. It is calculated as in Equation 7:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^2} \tag{7}$$

4. DATA AND APPLICATION

Observations of the gold variable used in the study are taken from the finance.yahoo website and closed values are used at the daily frequency between 01/01/2010 – 01/04/2023. The reason why this date range was chosen in the study includes the increase in gold prices in 2010 and the volatility that occurred in the global crisis in 2011. At the same time, the volatility in the price of gold was added to the model in the economic recession in the Covid-19 outbreak in 2019 and beyond. The polynominol kernel prediction process takes a lot of time. This data range has been studied as there is a timeout when working with more observations. The ARMA models are examined and ARMA (3,3) is determined before moving on to the GARCH models. The ARMA (3,3) prediction model is tested with ARCH-LM and rejected the basic hypothesis that the ARCH effect did not exist. As a result, it is concluded that it is appropriate to examine the GARCH models for the return series and the coefficient constraints of the models are examined.

The change in the gold price series variable over time is included in Figure 1.

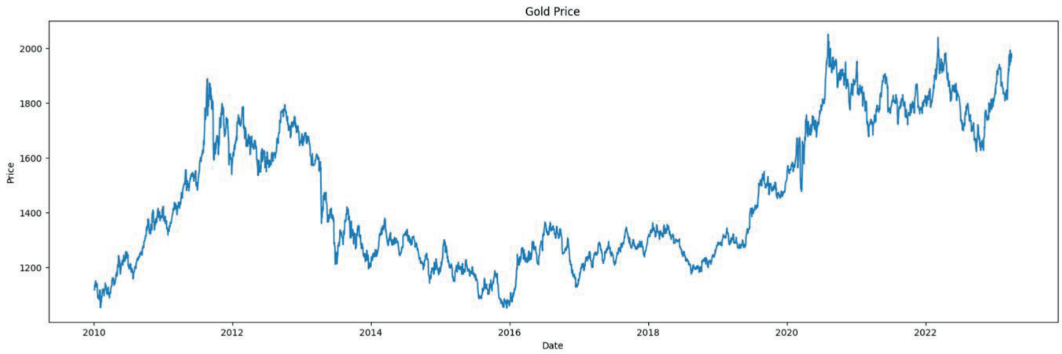

Figure 1. Gold Price Data

When Figure 1 is examined, it is seen that the series has a trend. Due to the fact that the series has fracture and extreme values, as well as the changing variance, the series is converted in the return series to avoid the problem. In the case of a changing variance problem, the term error indicates that variances are related to the error terms of past periods. The formula for the return series of the index is reported in Equation 8:

$$y_t = \ln\left(\frac{p_t}{p_{t-1}}\right) * 100 \quad (8)$$

The Equation in 8 y_t refers to the return of gold, refers to the price of gold. Figure 2 contains the graph of the gold return series.

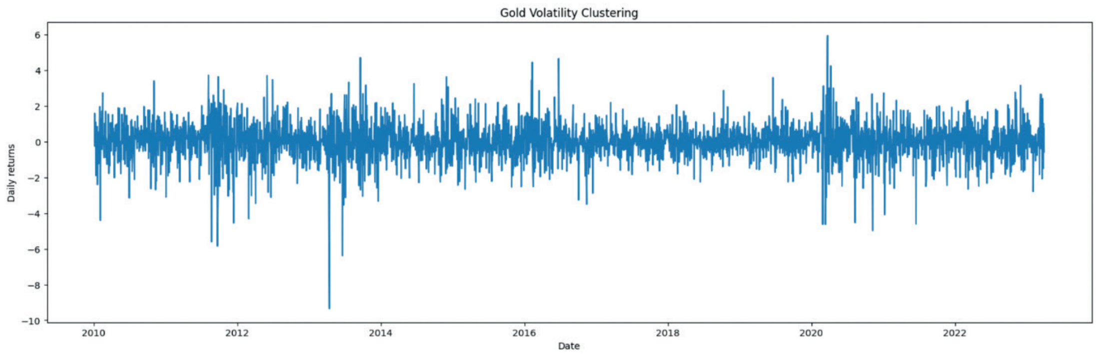

Figure 2. Gold Return Series

When Figure 2 is examined, it is seen that the series fluctuates around zero and has volatility clusters. Large shocks follow large shocks, while small shocks follow small shocks. Descriptive statistics for the gold price and return series are included in Table 2.

Table 2. Descriptive Statistics for the Gold Price and Return Series

	Price	**Return**
Number of observations	3424	3423
Mean	1460.444	0.01653
Median	1350.300	0.0269
Maximum	2069.400	5.7754
Minimum	1049.600	-9.8105
Std. Dev.	255.1708	1.0113
ADF	-1.4602	-59.8107
	(0.553)	(0.000)*
Skewness	0.4265	-0.5942
Kurtosis	1.7893	9.2772
Jarque-Bera	312.9352	5821.326
	(0.000)*	(0.000)*

Note: * indicates the rejection of the null hypothesis that the series is unit root according to 5% for the ADF test. Indicate the rejection of the null hypothesis that the distribution is normal according to 5% for the Jarque-Bera test.

The study is carried out with 3423 observations taking the return of the gold price, which is 3424 observations used in the study. ADF unit root test is applied to the gold price data and return data. While the price series has unit root, it is seen that the return series becomes stationary when the difference is taken. The Jarque-Bera test does not appear to have normal distribution for both the price series and the return series. 252 observations of 3423 observations are determined as test data, while the remaining observations are determined as data set training. As a model performance criterion, the RMSE error statistical criterion is taken into account.

Figure 3 shows the graphical output for ARCH model estimate.

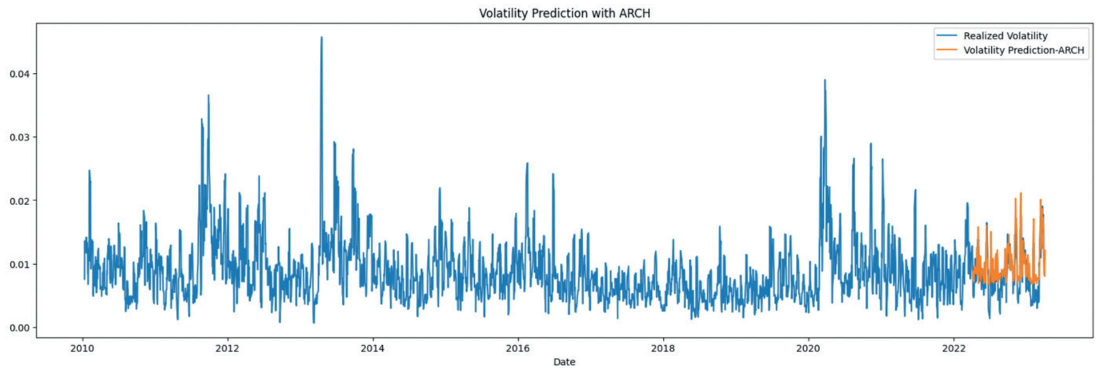

Figure 3. Comparison of the ARCH Model Volatility Estimate with the Actual Values

Figure 3 reveals that the ARCH model estimate does not exactly match the actual value. The forecast shows higher volatility than actual values.

Figure 4 shows the graphical output for the GARCH model estimate.

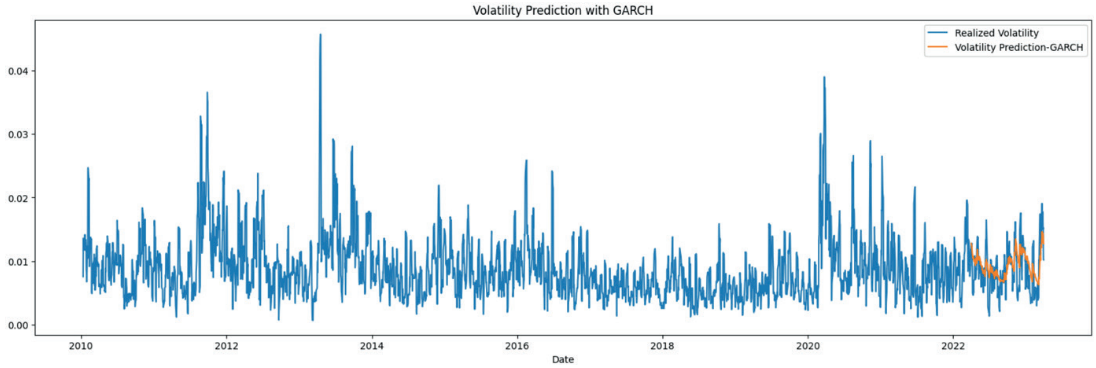

Figure 4. Comparison of the GARCH Model Volatility Estimate with the Actual Values

According to Figure 4, it is seen that the GARCH estimate does not overlap with the actual values and the estimate is insufficient.

Figure 5 shows the graphic output of the GJR-GARCH model estimate.

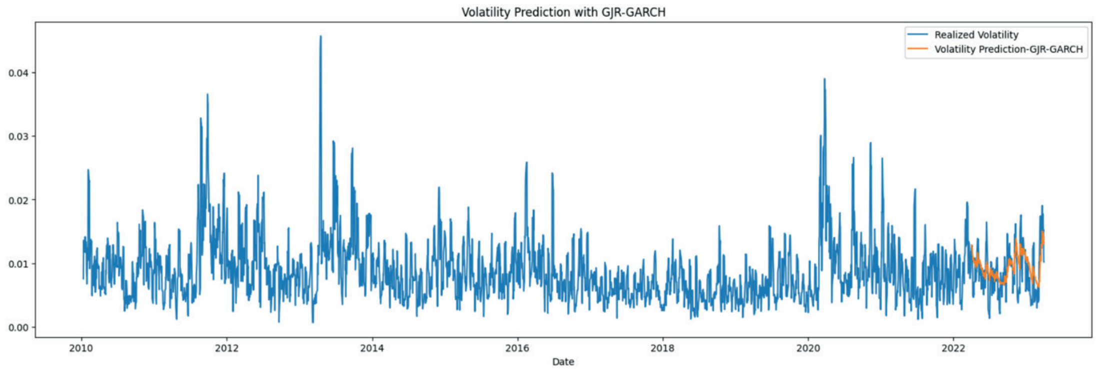

Figure 5. Comparison of the GJR-GARCH Model Volatility Estimate with the Actual Values

Once Figure 5 is taken into consideration, it could be seen that the actual volatility values and the GJR-GARCH model estimate do not overlap. This estimate appears to be insufficient to capture endpoints in volatility

Figure 6 shows the graphic output of the E-GARCH model.

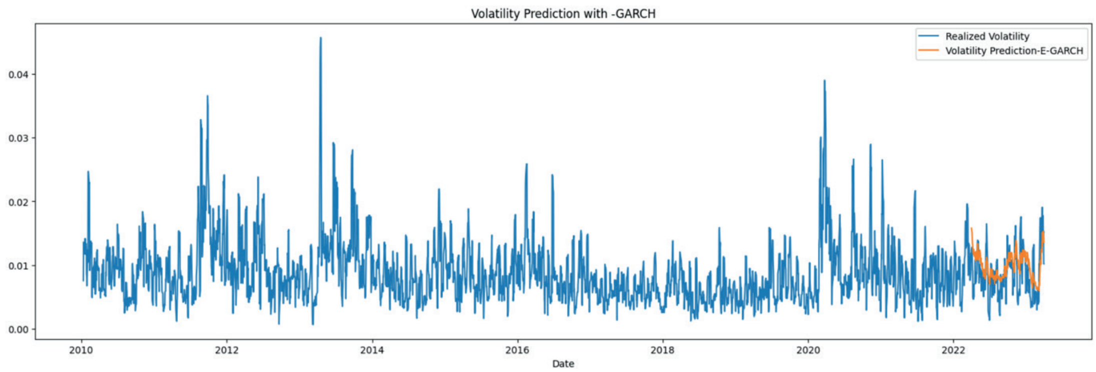

Figure 6. Comparison of E-GARCH Model Volatility Estimate with the Actual Values

Figure 6 suggests that E-GARCH cannot capture the actual volatility values with the estimate. This estimate method appears to have failed to catch endpoints.

Figure 7 shows the estimate graph for the linear SVR-GARCH model.

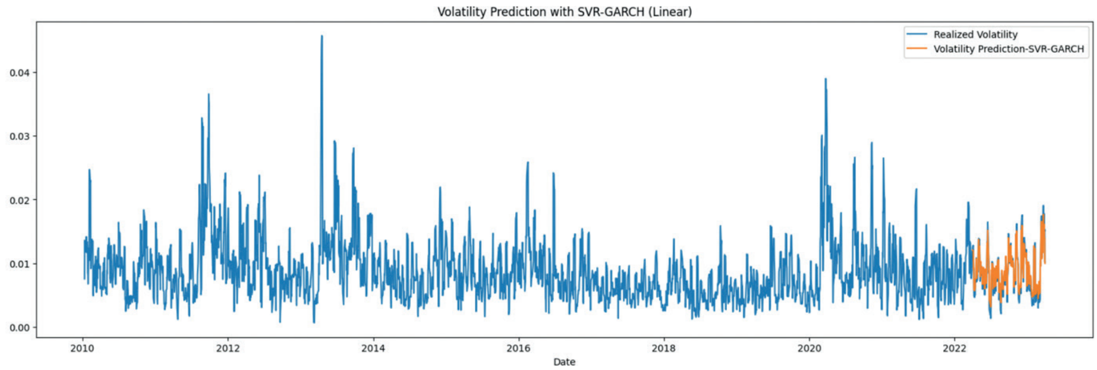

Figure 7. Comparison of Linear SVR-GARCH Model Volatility Prediction with the Actual Values

When Figure 7 is examined, it is seen that the linear SVR-GARCH model estimate shows very close estimate to actual values.

Figure 8 shows the estimate output of the RBF SVR-GARCH model.

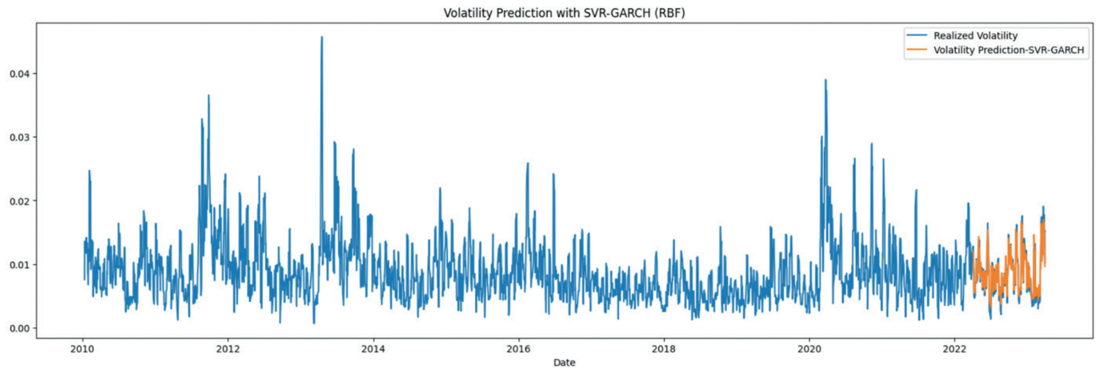

Figure 8. Comparison of RBF SVR-GARCH Model Volatility Estimate with the Actual Values

When Figure 8 is examined, it is seen that the RBF SVR-GARCH estimate overlaps with real value and is good at capturing endpoints. The forecast appears to have performed successfully in capturing the endpoints.

Figure 9 shows the graphic output for the Polynomial SVR-GARCH model estimate

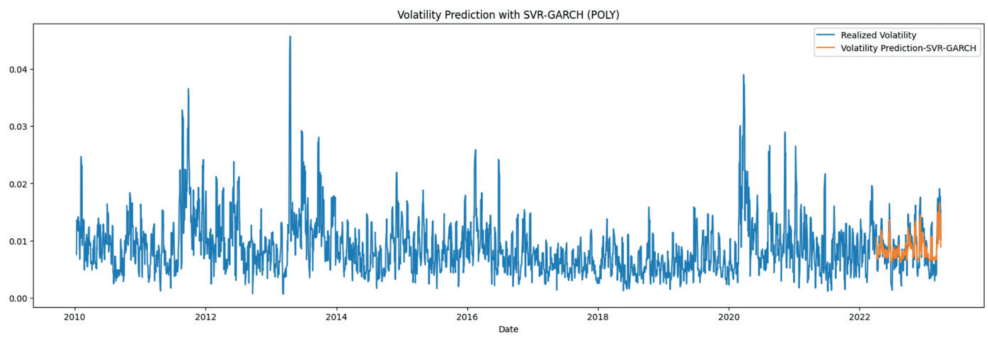

Figure 9. Comparison of Polynomial SVR-GARCH Model Volatility Prediction with the Actual Values

When Figure 9 is examined, it is seen that the Polynomial SVR-GARCH estimate not good at capturing real values and makes insufficient estimate.

The graphs of the prediction models can be interpreted, but the results we can rely on in the study are the model performance criteria. In this study, RMSE will be considered as the model performance criterion. Table 3 contains RMSE values as a model performance measure.

Table 3. Model RMSE Values

Prediction Method	RMSE	Model Performance Rank
ARCH	0.0913	7
GARCH	0.0879	4
GJR-GARCH	0.0880	5
E-GARCH	0.0904	6
Linear SVR-GARCH	0.0008	2
RBF SVR-GARCH	0.0007	1
Polinomial SVR-GARCH	0.0018	3

The RMSE values in Table 3 show that the models with the small statistics are hybrid models. The RMSE value of ARCH and GARCH models is 0.0913 and 0.0879, respectively. These models ignore the asymmetrical effect. GJR-GARCH has been proposed to solve the asymmetry problem in the financial series, and the RMSE value is 0.0880. GARCH and E-GARCH models feature short memory. This feature does not comply with long-term estimates. For this reason, the training set is kept smaller than the test set and the RMSE value is 0.0904. The SVR-GARCH model shows that the predictive performance is better, with less error statistics than GARCH models. SVR-GARCH models appear to give the RBF kernel with the smallest error statistics value 0.0007 due to the compatibility of the kernel to the data set.

5. CONCLUSION

Gold is the precious metal that people have used for centuries as a means of exchange and investment. Investors are curious about the future movements of gold prices because of being is a valuable mine. Investors want to know the volatility in order not to suffer losses and to make a maximum profit in their investments. Even countries try to understand the volatility in financial prices when making a decision to invest in each other. For this reason, it is important to determine the appropriate time series model for researchers.

In this study, the gold price series is obtained through yahoofinance.com and daily closing values are used. Traditional estimation methods of GARCH and hybrid SVR-GARCH Linear, RBF, Polinom kernels are used to take into account the nonlinear structure with time-varying volatility. When the graphs in the estimates are examined, the prediction effectiveness of GARCH models appears to be low. SVR-GARCH models, it seems more effective in predictions and closer to catching volatility. Considering the RMSE values determined as the model performance metric, it appears that the graphics are accurately related to the output and hybrid models have less error statistics. As a result of the study, it appears that the minimum error statistics value is belong to the hybrid model RBF kernel SVR-GARCH.

According to this study, considering the extreme volatility of the financial time series and its non-linear structure, it is concluded that hybrid models will be more accurate to prefer than traditional methods. In the study, ARCH, GARCH, GJR-GARCH, E-GARCH and SVR-GARCH models are predicted and compared. The SVR-GARCH hybrid model is created in three different prediction models as Linear, RBF and Polynomial kernels. When RMSE statistics and graphics are examined, it is seen that SVR-GARCH models have the best performance which is estimated by three different kernels. It is obvious that SVR-GARCH captures the volatility clusters in the graphics better. Therefore, it is suggested that SVR-GARCH hybrid models can be used in financial time series estimates.

Funding

This research did not receive any specific grant from the public, commercial, or not-for-profit funding agencies.

Submission Declaration Statement

We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere.

REFERENCES

- ALBERG, D., SHALIT, H., VE YOSEF, R. (2008). Estimating Stock Market Volatility Using Asymmetric GARCH Models. Applied Financial Economics, 18(15), 1201-1208.
- BEZERRA, P. C. S., & ALBUQUERQUE, P. H. M. (2017). Volatility forecasting via SVR–GARCH with mixture of Gaussian kernels. *Computational Management Science*, 14, 179-196.
- BILDIRICI, M., & ERSIN, O. O. (2012). Support Vector Machine GARCH Model in Modelling Conditional Volatility and an Application to Turkish Financial Markets. In *13th International Conference on Econometrics, Operations Research, and Statistic,* 24-26.
- BIRAU, R., TRIVEDI, J., & ANTONESCU, M. (2015). Modeling S&P Bombay Stock Exchange BANKEX Index Volatility Patterns Using GARCH Model. Procedia Economics and Finance, 32, 520-525.
- BOLLERSLEV, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3), 307-327.
- CIHANGIR, Ç. K., & UGURLU, E. (2017). Volatility in Gold Market: Model Recommendation for Turkey. *İşletme Araştırmaları Dergisi*, 9(3), 284-299.
- CAGLAYAN, E., & DAYIOGLU, T. (2009). Döviz Kuru Getiri Volatilitesinin Koşullu Değişen Varyans Modelleri ile Öngörüsü. *Istanbul University Econometrics and Statistics e-Journal*, 9, 1-16.
- ENGLE, R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of The Variance of UK Inflation. *Econometrica,* 50, 987–1008.
- ENGLE, R. F. (1993), “Statistical Models for Financial Volatility”, Financial Analyst Journal, 49(1), 72-78.
- ENGLE, R. (2001). GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics. *Journal of Economic Perspectives*, 15(4), 157-168.
- ENGLE, R. F., & SOKALSKA, M. E. (2012). Forecasting Intraday Volatility in the Us Equity Market. Multiplicative Component Garch. *Journal of Financial Econometrics*, 10(1), 54-83.
- GLOSTEN, L. R., JAGANNATHAN, R., & RUNKLE, D. E. (1993). On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The journal of finance*, 48(5), 1779-1801.
- GOYARI, P., & JENA, P. K. (2010). Is the Recent Global Financial Crisis Causes Volatility in Gold and Crude Oil Futures Market in India?. In *International Conference on Applied Economics–ICOAE*, 353-361.
- GURSOY, M., & BALABAN, M. (2014). Hisse Senedi Getirilerindeki Volatilitenin Tahminlenmesinde Destek Vektör Makinelerine Dayalı Garch Modellerinin Kullanımı. *Kafkas Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 5(8), 167-186.
- KARABACAK, M., MECIK, O., & ERHAN, G. E. N.C. (2014). Koşullu Değişen Varyans Modelleri ile BİST 100 Endeks Getirisi ve Altın Getiri Serisi Volatilitesinin Tahmini. *Uluslararası Alanya İşletme Fakültesi Dergisi*, 6(1), 79-90.
- KATSIAMPA, P. (2017). Volatility Estimation for Bitcoin: A Comparison of GARCH Models, Economics Letters, 158, 3-6.
- NELSON, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica: Journal of The Econometric Society*, 59(2), 347-370.
- OU, P., & WANG, H. (2010). Financial Volatility Forecasting by Least Square Support Vector Machine Based on GARCH, EGARCH and GJR Models: Evidence from ASEAN Stock Markets. *International Journal of Economics and Finance*, 2(1), 51-64.

- PENG, Y., ALBUQUERQUE, P. H. M., DE SÁ, J. M. C., PADULA, A. J. A., & MONTENEGRO, M. R. (2018). The Best of Two Worlds: Forecasting High Frequency Volatility for Cryptocurrencies and Traditional Currencies with Support Vector Regression. *Expert Systems with Applications*, 97, 177-192.

- PÉREZ-CRUZ, F., AFONSO-RODRIGUEZ, J. A., & GINER, J. (2003). Estimating GARCH Models Using Support Vector Machines. *Quantitative Finance*, 3(3), 163.

- SUN, H., & YU, B. (2020). Forecasting Financial Returns Volatility: A GARCH-SVR Model. *Computational Economics*, 55, 451-471.

- TAY, F. E., & CAO, L. (2001). Application of Support Vector Machines in Financial Time Series Forecasting. *Omega*, 29(4), 309-317.

- TAYLOR, S. (1986). Modeling Financial Time Series John Wiley & Sons. *Great Britain*.

- THOMAS, S., & MITCHELL, H. (2005). GARCH Modeling of High-Frequency Volatility in Australia's National Electricity Market. *System*, 1-39.

- WANG, Y., & WU, C. (2012). Forecasting Energy Market Volatility Using GARCH Models: Can Multivariate Models Beat Univariate Models?. *Energy Economics*, 34(6), 2167-2181.

- WANG, B., HUANG, H., & WANG, X. (2013). A Support Vector Machine Based MSM Model for Financial Short-Term Volatility Forecasting. *Neural Computing and Applications*, *22*, 21-28.

JOURNAL OF APPLIED MICROECONOMETRICS

JAME

HOLISTENCE
publications