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: 4
Issue: 1
2024

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

International Peer-Reviewed and Open Access Electronic Journal

Volume: 4

Issue: 1

Year: 2024

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

HOLISTENCE
publications

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

"This page is left blank for typesetting"

HOLISTENCE
publications

CONTENTS

"This page is left blank for typesetting"

JAME
E-ISSN:2791-7401
Volume : 4, Issue 1, 2024
URL: https://journals.gen.tr/jame

Received: 2024-05-14
Accepted: 2024-06-02

RESEARCH ARTICLE

Examining the relationship between financial ratios and stock returns: An application on BIST 30 index

Özgün Şanlı

Lec. Dr., Tekirdağ Namık Kemal University, Saray Vocational School, Türkiye, email: osanli@nku.edu.tr

Abstract

Investors trading in capital markets aim to maximize the returns they will obtain from this market. For this reason, determining the factors affecting stock returns is important for investors. The aim of this study is to examine the relationship between financial ratios and stock returns of companies that are listed on the BIST 30 Index as of 2024 and traded on the stock exchange uninterruptedly between the 2016Q2-2023Q4 periods. The financial ratios used in the research include the current ratio, return on equity ratio, asset turnover ratio, inventory turnover ratio, debt/equity ratio, and debt/asset ratio. Stock returns are measured by the rate of return. The relationship between the return rates of stocks of companies listed on the BIST 30 index and the financial ratios of these companies will be examined through the panel data analysis method. In the analysis results; According to the analysis results, the relationship between the current ratio and inventory turnover ratios and the return rate of stocks is significant and negative. The relationship between return on equity ratio, asset turnover ratio and debt/equity ratio and stock returns is significant and positive. The relationship between debt/asset ratio and return rate is meaningless.

Keywords: Financial Ratios, Stock Returns, Panel Data Analysis, BIST 30 Index, Driscoll-Kraay Robust Standard Estimator

JEL codes: C10, C50, G11, G20,

Citation: ŞANLI, Ö. (2024). Examining the relationship between financial ratios and stock returns: An application on BIST 30 index. *Journal of Applied Microeconometrics (JAME)*. 4(1),1-11, DOI: 10.53753/jame.2424

Corresponding Author:
Özgün Şanlı
E-mail: osanli@nku.edu.tr

1. INTRODUCTION

Investors trading in capital markets aim to maximize profit by using their funds effectively. To achieve this goal, they can invest in some capital market instruments such as stocks. Investors may want to gain two types of profits when investing in stocks. These include dividends and capital gains. Dividend is the profit share that investors receive from the businesses in which they are partners, proportion to their capital. The positive difference between the selling price and the buying price of any financial asset is called a capital gainp. Investors in the market can increase capital gains by investing in stocks with appreciation potential. In this case, factors affecting the price of stocks can be significant for investors.

When the financial literature is examined, it is seen that there are many factors that have significant effects on stock returns. The financial performance of businesses whose stocks are traded in capital markets may affect the market price of their stocks. The market price of shares of a business with good financial performance is likely to increase. Additionally, current economic conditions in the country may also affect the market prices of stocks. For example, an increase in deposit interest rates may cause funds in capital markets to turn towards money markets. This may reduce the market price of the stocks. Political factors in the country can also affect the market price of stocks. In general, if there are positive expectations about the future of the country, the market price of stocks may increase. By taking such factors into consideration, investors can increase their profits from capital markets.

The aim of this study is to determine the financial ratios that are effective on stock returns. For this purpose, the data of 23 different companies registered in the BIST 30 Index as of 2024 and continuously traded on the stock exchange between 2016Q2-2023Q4 periods are analysed. BIST 30 Index consists of 30 firms with the highest trading volume and market capitalisation traded in Borsa Istanbul. Since it is thought that the results of the analyses to be made with the data of the companies in this index will be more meaningful, it is preferred in the study. Since all data of the companies traded in BIST 50 and BIST 100 Indices are not available, these indices are not preferred in the study. In the study, firstly, a literature review consisting of studies examining the relationship between stock returns and financial ratios is presented. Then, the scope of the study, hypotheses, method of analysis and research findings are presented. Finally, in the conclusion section, the results and recommendations obtained as a result of the research are given.

2. LITERATURE REVIEW

When the literature was examined, it was seen that there were many studies examining the relationship between financial ratios and stock returns. One of the first studies on the subject was conducted by Senchack and Martin in 1987. Senchack and Martin (1987) analysed the relationship between financial ratios and stock returns using the data of 450 firms listed on AMEX and NYSE. According to the results of the analysis, financial ratios have an effect on stock returns.

Martikainen (1989) analysed the relationship between 12 different financial ratios and stock prices. According to the results of the study, profitability and capital structure ratios have significant effects on stock returns.

Fama and French (1992) examined the relationship between stock returns and financial ratios using data on firms traded on NYSE, AMEX and NASDAQ stock exchanges. According to the results of the study, market to book ratio is significant in explaining stock returns.

Lev and Thiagarajian (1993) examined the relationship between 12 different financial performance measures and stock returns. Some variables such as inventories, gross sales profit, accounts receivable were used in the study. According to the results of the study, there are high correlations between financial performance measures and stock returns.

Haugen and Baker (1996); examined the data of American companies traded in the Russell 3000 Index for the period 1979-1993. According to the research results, there are positive and significant relationships between stock returns and profitability.

Dhatt et al. (1999) investigated the relationship between financial performance and stock returns of companies listed on the Korea Stock Exchange. The research used data of these companies for the years 1982-1992. Dhatt et al. found significant relationships between Book value/market value ratio, debt/equity ratio and total sales/market value ratio and stock return.

Omran and Ragap (2004) examined the data of a total of 46 companies operating in Egypt for the period 1996-2000. In the research; the relationship between liquidity, leverage, activity and profitability ratios and stock returns was examined. Omran and Ragap argued that there are non-linear relationships between some financial ratios and stock returns.

Kalaycı and Karataş (2005) investigated the relationship between stock returns and some financial ratios by examining the data of firm traded on the Borsa Istanbul. Six-month financial statements for the period 1996-1997 were used in the research. According to the results obtained in the research; significant relationships were found between business profitability, stock market performance and productivity rates and stock returns.

Alexakis et al. (2010) examined data from 47 different companies registered on the Athens Stock Market. In the research, the financial performances of these companies and the returns of their stocks were analyzed. In the research, companies' data for the years 1993-2006 were examined. According to the research results; the relationships between profitability, asset turnover, price/earnings, market value/book value and current ratio and stock return are significant.

Kheradya et al. (2011) analyzed the data of 960 companies registered on the Malaysian Stock Exchange between 2000 and 2009 to examine relationship between firm value, price/earnings and market value/book value ratios and stock returns. Analysis results; it has been shown that the relationship between firm value, price/earnings and market value/book value ratios and stock returns is significant.

Arkan (2016) investigated the financial ratios that are considered to have an impact on stock returns through data obtained from 15 companies operating in three different sectors registered in the Kuwait Stock Exchange between 2005 and 2014. In his analyses, Arkan (2016) concluded that the relationship between firms' financial performance and stock returns differs according to the line of business in which the firms operate.

Allozi and Obeidat (2016) analysed relationship between stock returns and financial ratios using data of 65 firms traded on the JSE between 2001 and 2011. Leverage and profitability ratios of companies were used as independent variables in the research. The results obtained by Allozi and Obeidat showed that the return on equity ratio can affect stock returns.

Sarı and Kırkık (2019) examined the relationship between some financial ratios and stock returns using data from 2006-2015 period of 20 companies registered in Borsa Istanbul and operating in the manufacturing sector. According to the analysis results, there are positive and significant relationships between stock returns and activity, liquidity and profitability ratios. Another result obtained from Sarı and Kırkık's research is that debt ratios do not affect stock returns.

Patin et al. (2020) using data from 1961 US companies for the period 2001-2015, it examined the relationship between stock return and total asset turnover ratio. According to the results obtained in the research, there are positive significant relationships between stock return and total asset turnover ratio.

Uyar and Sarak (2020) investigated whether some financial ratios have an effect on stock returns, using data from the period 2008-2018 of a total of 81 companies traded on Borsa Istanbul and the London Stock Exchange. According to the research results, the ratio that has the highest power to explain the returns of stocks traded on Borsa Istanbul is the current asset turnover rate, and the rate that has the highest power to explain the returns of stocks traded on the London Stock Exchange is the return on equity ratio.

Apan and Öztel (2021) aimed to define relationship between stock return and financial performance of banks by using 2015-2019 data of deposit banks registered in the BIST-Bank Index. According to the results obtained by

Apan and Öztel, financial performance of banks does not affect stock returns.

Tekin and Bastak (2022) used data from the 2010-2018 period of companies traded in the BIST 100 Index to determine the internal factors affecting stock returns. According to the results of the research, negative relationships were found between leverage ratio, liquidity ratio, current asset turnover rate and stock return. In addition, in the analysis results; Positive relationships were found between current ratio, return on equity, asset turnover rate and stock return.

3. RESEARCH METHODOLOGY

The aim of this study is to determine some financial ratios that are thought to have an impact on the stock returns of enterprises. In this study, it is aimed to analyse the relationship between some financial ratios of the companies registered in the BIST 30 Index as of 2024 and continuously traded on the stock exchange between 2016Q2-2023Q4 periods and the stock returns of these enterprises. Panel data analysis method was used in the study. In order to create a balanced panel data set, the period 2016Q2-2023Q4 was selected in the study. In case of going beyond the period 2016Q2-2023Q4, there may be missing data in some variables Additionally, data from companies operating in the financial sector were not included in the research. In this context, the data set of the research consists of 23 companies. In the research, from the financial ratios of the companies in question current ratio, return on equity, debt/ equity ratio, debt/asset ratio, asset turnover ratio and inventory turnover ratios were used. Stock returns are measured by the rate of return. The data used in the research was obtained from www.fintables.com and www.finnet2000.com .

The hypotheses of this study, which was conducted to examine the relationship between financial ratios and stock returns, are as follows:

H_1 : There are significant relationship between current ratios and stock returns of the companies registered in the BIST 30 index

H_2 : There are significant relationship between return on equity ratios and stock returns of the companies registered in the BIST 30 index

H_3 : There are significant relationship between debt/equity ratios and stock returns of the companies registered in the BIST 30 index

H_4 : There are significant relationship between debt/asset ratios and stock returns of the companies registered in the BIST 30 index

H_5 : There are significant relationship between asset turnover ratios and stock returns of the companies registered in the BIST 30 index

H6 : There are significant relationship between inventory turnover ratios ratios and stock returns of the companies registered in the BIST 30 index.

The variables to be used for testing the research hypotheses are as follows:

Rate of Return (RoR): The dependent variable of the research is the rate of return of stocks. In the calculation of the rates of return, the following formula was utilised (Hallerbach, 2005; Ünlü et al., 2009):

$$RoR_t = \ln(P_t / P_{t-1}) \quad (1)$$

RoR_t = Rate of return

ln = Natural Logarithm

P = Stocks Prices

t = time

Current Ratio (CR): The first independent variable of the research is the current ratio. The ratio is used to measure the ability of the business to pay its short-term debts on time. Companies need to keep the current ratio high in order to pay their short-term debts on time. The ratio is calculated as in the formula (Daryanto and Nurfadilah, 2018:13):

$$CR_{t-1} = (\text{Current Assets}_{t-1} / \text{Current Liabilities}_{t-1}) \quad (2)$$

Return on Equity (ROE): Return on equity is a ratio that shows the extent to which the capital contributed by the shareholders to the company is used effectively. The expectations of company shareholders may be in favour of a high return on equity ratio. ROE formula (Daryanto and Nurfadilah, 2018:13):

$$ROE_{t-1} = (\text{Net Income}_{t-1} / \text{Shareholders' Equity}_{t-1}) \quad (3)$$

Debt /Equity Ratio (D/E): The D/E ratio, which shows the ability of companies to pay their debts, also shows the business risk. D/E ratio formula (Colline, 2022:81):

$$D/E_{t-1} = (\text{Total Debt}_{t-1} / \text{Total Equity}_{t-1}) \quad (4)$$

Debt / Asset Ratio (D/A): This ratio is a ratio that shows how much of the assets owned by the business are purchased with debts. A high ratio may increase the risk of the business not being able to pay its debts. D/A ratio formula (Doğan, 2013:181):

$$D/A_{t-1} = (\text{Total Debt}_{t-1} / \text{Total Assets}_{t-1}) \quad (5)$$

Asset Turnover Ratio (ATO): It is calculated as the ratio of the sales revenue realised by an enterprise in a certain period to the value of the enterprise assets in the same period. A high ratio indicates that the performance of the enterprise is good. ATO ratio formula (Utami, 2017:27):

$$ATO_{t-1} = (\text{Sales}_{t-1} / \text{Total Assets}_{t-1}) \quad (6)$$

Inventory Turnover Ratio (ITO): It is a performance measure that shows how many times the inventories owned by an enterprise are sold and renewed within a period. Shareholders may want this ratio to be above the sector average. ITO ratio formula (Daryanto and Nurfadilah, 2018:13):

$$ITO_{t-1} = (\text{Cost of Goods Sold}_{t-1} / \text{Average Inventory}_{t-1}) \quad (7)$$

The dependent variable of the research is the RoR. The independent variables are CR, ROE, D/E, D/A, ATO and ITO. Panel data analysis method will be used to obtain research findings.

The regression model constructed to analyse the relationship between the stock returns of the companies and the independent variables is as follows:

$$RoR_{it} = \alpha + \beta_1 CR_{it-1} + \beta_2 ROE_{it-1} + \beta_3 D/E_{it-1} + \beta_4 D/A_{it-1} + \beta_5 ATO_{it-1} + \beta_6 ITO_{it-1} + \varepsilon_{it} \quad (8)$$

i = 1,......,23

t = 1,......,30

The i in the model represent the cross-sectional units, the t represent the time series in the panel data, α is the constant term, β is the coefficients of the independent variables, and ε is the error term in terms of periods and units.

The data used in the research include time series and more than one horizontal cross-section unit. For this reason, it can be said that the data set has the characteristics of panel data. Panel data is a type of data in which different data belonging to more than one unit are presented together for different periods. In other words, panel data consists of *n* number of units and *t* number of observations corresponding to each unit (Tatoğlu , 2016: 2). In panel data, if there is a time series of equal length for each cross-sectional unit, there is a balanced panel; if there is no time series of equal length for each cross-sectional unit, there is an unbalanced panel (Çetin and Ecevit, 2010: 172).

It can be said that there are three different panel data models. These models can be listed as classical model, fixed effects model and random effects model. In the classical model, both constant and slope parameters are assumed to be homogeneous across units and time. In the fixed effects model, while the slope parameters are the same for all cross-sectional units, the constant term takes a different value for each cross-sectional unit. In other words, in the fixed effects model, unit effects are transferred to the model through the constant term. In the random effects model, differences between units are expressed with error terms. In the random effects model, the error term consists of two different components. These are residual errors and unit errors (Tatoğlu , 2018: 37-103). The features, assumptions and estimation methods of each panel data model are different. For this reason, choosing the right model and estimation method is important for the reliability of the analysis results.

4.RESEARCH FINDINGS

In this section of the study, the relationship between some financial ratios (CR, ROE, D/E, D/A, ATO, ITO) of 23 different companies included in the study and the stock returns of these companies will be analysed and the results of the analysis will be interpreted. Under this heading, first descriptive statistics regarding the variables will be included. Then, the correlation matrix that reveals the correlation relationship between the independent and dependent variables of the regression model will be presented. Then, the findings regarding the panel data analysis applied to the research data set will be included. Descriptive statistics of CR, ROE, D/E, D/A, ATO, ITO variables are as in Table 1.

Table 1. Descriptive Statistics

Variable	Observation	Average	Standard deviation	Minimum	Maximum
RoR	690	0.240886	0.152041	-0.230258	0.672085
CR	690	1.820696	2.009182	0.35	14.91
ROE	690	25.42965	31.96496	-85.34	199.49
D/E	690	58.19362	20.39296	7.82	93.06
D/A	690	29.83046	18.38473	0	70.6
ATO	690	1.015014	0.9768557	0.04	8.01
ITO	690	17.97203	32.9553	1.39	222.38

Table 1, shows that the number of observations is 690 for all variables. This shows that the panel data set is balanced. The table shows that the standard deviations of the CR, ROE and ITO variables are higher than their average values. Based on this result, it can be said that the companies included in the research differ significantly in terms of CR, ROE and ITO variables.

After the descriptive statistics of CR, ROE, D/E, D/A, ATO, ITO variables, the correlation matrix for these variables will be presented. The correlation relationship between variables can provide information about the multicollinearity problem. The correlation relationship between CR, ROE, D/E, D/A, ATO, ITO variables and the significance level of these correlation relationships are presented in Table 2. The correlation coefficient between the independent variables of the study can provide information about the problem of multicollinearity.

Table 2. Correlation Matrix Between Variables and VIF Value

Variables	RoR	CR	ROE	D/E	D/A	ATO	ITO	VIF
RoR	1							-
CR	-0.2334*	1						1.87
ROE	0.4068*	-0.0099	1					1.27
D/E	0.3104*	-0.6791*	0.0343	1				2.74
D/A	0.0740**	-0.4460*	-0.1155*	0.6586*	1			2.03
ATO	0.3891*	-0.1269*	0.4477*	0.1432*	-0.0762**	1		1.32
ITO	0.2497*	-0.1085*	-0.0794**	0.2395*	0.3658*	-0.0220	1	1.16

Note: * 1%, ** 5%, *** 10% indicate the significance level.

When the correlation coefficients between the independent variables to be used in the analysis are examined, it is seen that they are below the critical value (0.8) recommended by Gujarati and Porter (2009). Since there is no high correlation between all variables, the multicollinearity problem is expected not to distort the results (Tuan and Borak, 2020: 388). In addition, VIF values of the independent variables used in the study are given in the Table 2. If the VIF value is greater than 10, it can be said that there is a multicollinearity problem between independent variables (Büyükuysal and Öz, 2016:111; Topaloğlu, 2018:294; Alkan and Abar, 2019:7; Shrestha, 2020:40). When the table is analysed, it is seen that the VIF values of the independent variables used in the research are less than 10. For this reason, it can be said that there is no multicollinearity problem between the variables.

In panel data analysis, it is very important to perform some diagnostic tests to identify the correct estimator. The first diagnostic test applied to the data set of the research is the F test. As a result of the F test, the test statistic was found to be 30.82 at the 0.0000 significance level. When this value is compared with F(22,661) degrees of freedom in the F distribution table, the H_0 hypothesis of the F test is rejected.

Another test that should be applied to the data set of the study in order to determine the correct panel data model is the LM test. The LM test applied to the data set of this research was found to be 1185.15 at the 0.0000 significance level. When this value is compared with the chi-square table, the H_0 hypothesis of the LM test is rejected.

As a result of both tests, the H_0 hypotheses of the said tests were rejected. Therefore, the Hausman test was needed to select the panel data model suitable for the research data set. The Hausman test applied to determine the appropriate panel data model for the data set of the study yielded a test statistic of 14.48 at a significance level of 0.0257. Considering the obtained test statistics and significance value, the H_0 hypothesis of the Hausman test was not accepted. As a result of the diagnostic tests applied to the data set, it was concluded that the fixed effects model was appropriate.

The fixed effects model has some assumptions. Assumptions need to be tested to determine the correct estimator. One of the assumptions of the fixed effects model is homoskedasticity. Modified Wald Test is used to test this assumption. As a result of the Modified Wald Test, the test statistic was 403.31 at the 0.0000 significance level. This result obtained as a result of the analysis shows that the H_0 hypothesis of the test is not accepted. Based on this, it can be said that the model to be used in the research is heteroskedastic.

Baltagi-Wu Locally Best Invariant Test was used to test the non-autocorrelation assumption. As a result of the test, Modified Bhargava et al. Durbin-Watson test statistic was found to be 0.3585 and Baltagi - Wu LBI statistic was 0.5054. According to this result, the research model contains autocorrelation. Pesaran's cross-sectional dependence test was used to test the assumption of inter-unit correlation. As a result of the test, the test statistic was 47.723 at the 0.0000 significance level. The test statistic is above the critical value stated by Pesaran (2004). According to this result, there is a correlation between units in the research model. The test results to determine the panel data estimator applied to the research data set are given in the Table 3.

Table 3. Analysis Results for Identifying the Correct Estimator

Tests for Model Selection	**Analysis Results**
F Test	30.82*
LM Test	1185.15*
Hausman Test	403.31**
Autocorrelation Test	
Modified Bhargava et al. Durbin -Watson test statistic	0.3585
Baltagi - Wu LBI test statistic	0.5054
Heteroscedasticity Test	
Modified Wald Test Statistics	403.31*
Cross Section Dependency Test	
Pesaran CD Test statistics	47,723*

Note: * 1%, ** 5%, *** 10% indicate the significance level.

According to the test results for model selection, the panel data model suitable for the research data set is the fixed effects model. According to the analysis results that tested the assumptions of the fixed effects model, the research model includes heteroscedasticity, autocorrelation and cross-sectional dependence. According to the analysis results, the Driscoll-Kraay robust standard estimator is suitable for analysis. Driscoll-Kraay robust standard estimator results are given in the Table 4.

Table 4. Regression Analysis Results

Independent variables	**Coefficient**	**Driscoll – Kraay Standard Errors**	**Probability Value**
CR	**-0.043575**	**0.0154034**	**0.008**
ROE	**0.0021111**	**0.0002538**	**0.000**
D/E	**0.0020991**	**0.0007063**	**0.006**
D/A	0.0012576	0.0008171	0.135
ATO	**0.0403132**	**0.0918383**	**0.000**
ITO	**-0.0007157**	**0.0003609**	**0.057**
Constant Term	0.0788147	0.0659417	0.242
R^2-value	0.3978	**Number of Observations**	690
F-value	0.0000	**Number of Companies**	23

According to the analysis results, the F-value was found to be 0.0000. This result shows that the applied regression model is significant. In the analysis results, the R^2 value was found to be 0.3978. According to the regression analysis results, the relationship between the independent variables CR, ROE, D/E, ATO, ITO and the dependent variable RoR is statistically significant. The relationship between the independent variables CR and ITO and dependent variables is negative. According to the results of the analyses, investors who invest in the stocks of enterprises with low current ratio and inventory turnover rate can earn high returns. The relationship between ROE, D/E, ATO and RoR is positive. According to the results of the analyses, the stock returns of enterprises with high return on equity, debt/equity and asset turnover ratios are also high. It is likely that the market price of the stocks of enterprises with high profitability is high. Because, high profitability can be perceived as a positive signal by investors. According to the net income approach, which is one of the capital structure policies, the higher the debt/equity ratio, the higher the firm value. The results obtained in the study are consistent with the net income approach. Additionally, an increase in the asset turnover rate may be perceived as a positive signal by capital markets. This may increase stock returns. According to the analysis results the H_1, H_2, H_3, H_5 and H_6 hypotheses of the study have not been rejected.

5. CONCLUSION

The aim of individuals trading in capital markets is to maximize their personal wealth. Individuals carry out buying and selling activities in financial markets for this purpose. Investors trading in the stock market can earn two types of profits. These are dividends and capital gains. Not all investors in the capital markets can control dividend earnings. However, it can increase capital gains by applying the right trading strategies. For this reason, factors affecting stock returns are important for investors.

In this study, which was conducted to examine the returns between stock returns and financial ratios, the data of companies registered in the BIST 30 Index as of 2024 and traded on the stock exchange without interruption between the periods of 2016Q2-2023Q4 were examined. Additionally, companies operating in the financial sector were not included in the data set of the research. In the research, the effects of CR, ROE, D/E and D/A ratio, ATO and ITO on stocks returns were examined. Panel data analysis method was used to examine the data of the companies included in the research. According to the analysis results applied to the research data set, the relationship between CO, ROE, D/A, ATO, ITO and rate of return is statistically significant. This result supports the results obtained by Haugen and Baker (1996), Dhatt et al. (1999), Omran and Ragap (2004), Kalaycı and Karataş (2005), Alexakis et al. (2010), Allozi and Obeidat (2016), Sarı and Kırkık (2019) and Tekin and Bastak (2022). The results obtained from the research show that financial ratios have an effect on stock returns. In subsequent studies, it can be examined whether the relationship between stock returns and financial ratios differs during the economic crisis. The relationship between liquidity ratios, operating ratios, borrowing ratios and profitability ratios of enterprises and stock returns can be analysed on enterprises operating in different sectors. Thus, it can be determined whether the relationship between these variables differs across sectors.

REFERENCES

- ALEXAKİS, C., PATRA, T., & POSHAKWALE, S. (2010). Predictability of stock returns using financial statement information: evidence on semi-strong efficiency of emerging Greek stock market. *Applied Financial Economics*, *20*(16), 1321-1326.

- ALKAN, Ö., & ABAR, H. (2020). Determination of factors influencing tobacco consumption in Turkey using categorical data analyses. *Archives of environmental & occupational health*, *75*(1), 27-35.

- AL-LOZI, N. M., & OBEIDAT, G. S. (2016). The relationship between the stock return and financial indicators (profitability, leverage): An empirical study on manufacturing companies listed in Amman Stock Exchange. *Journal of Social Sciences (Coes&Rj-Jss)*, *5*(3), 408-424.

- APAN, M., & ÖZTEL, A. (2021). Pay Senedi Getirisi İle Finansal Performans Arasındaki İlişki: Bulanık Ahp-Gia Yöntemleri Kullanılarak Bist-Banka Endeksi'nde Ampirik Bir Araştırma. *Finans Ekonomi ve Sosyal Araştırmalar Dergisi*, *6*(4), 917-932.

- ARKAN, T. (2016). The importance of financial ratios in predicting stock price trends: A case study in emerging markets. *Finanse, Rynki Finansowe, Ubezpieczenia*, (79), 13-26.

- BÜYÜKUYSAL, M. Ç., & ÖZ, İ. İ. (2016). Çoklu doğrusal bağıntı varlığında en küçük karelere alternatif yaklaşım: Ridge regresyon. *Düzce Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi*, *6*(2), 110-114.

- COLLINE, F. (2022). The Mediating Effect of Debt Equity Ratio on The Effect of current ratio, return on equity and total asset turnover on price to book value. *Jurnal Keuangan dan Perbankan*, *26*(1), 75-90.

- ÇETİN, M. & ECEVİT, E., (2010), Sağlık Harcamalarının Ekonomik Büyüme Üzerindeki Etkisi: OECD Ülkeleri Üzerine Bir Panel Regresyon Analizi, *Doğuş Üniversitesi Dergisi*, 11(2), (166-182).

- DARYANTO, W. M., & NURFADILAH, D. (2018). Financial performance analysis before and after the decline in oil production: Case study in Indonesian oil and gas industry. *International Journal of Engineering & Technology*, *7*(3.21), 10-15.

- DHATT, M. S., KIM, Y. H., & MUKHERJİ, S. (1999). Relations between stock returns and fundamental variables: Evidence from a segmented market. *Asia-Pacific Financial Markets*, *6*, 221-233.

- DOĞAN, M. (2013). Katılım ve geleneksel bankaların finansal performanslarının karşılaştırılması: Türkiye örneği. *Muhasebe ve Finansman Dergisi*, (58), 175-188.

- FAMA, E. F., & FRENCH, K. R. (1992). The cross‐section of expected stock returns. *the Journal of Finance*, *47*(2), 427-465.

- GUJARATİ, D.N. & PORTER, D.C., (2009), Basic Econometrics, McGraw-Hill

- HALLERBACH, W. G., (2005), Holding Period Return-Risk Modelling: The Importance of Dividends, *Estudios De Economia Aplicada,* 23(1), 45-65.

- HAUGEN, R. A., & BAKER, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of financial economics*, *41*(3), 401-439.

- KALAYCI, Ş., & KARATAŞ, A. (2005). Hisse senedi getirileri ve finansal oranlar ilişkisi: İMKB'de bir temel analiz araştırması. *Muhasebe ve Finansman Dergisi*, (27), 146-158.

- KHERADYAR, S., İBRAHİM, I., & NOR, F. M. (2011). Stock return predictability with financial ratios. *International Journal of Trade, Economics and Finance*, *2*(5), 391.

- KIRKIK, Ş., & SARI, S. (2018). Borsa İstanbul'da İşlem Gören Hisse Senetlerinin Muhasebe Verileriyle İlişkilendirilmesi: Bir Sektör Uygulaması. *Izmir Democracy University Social Sciences Journal*, *1*(1), 1-15.

- LEV, B., & THİAGARAJAN, S. R. (1993). Fundamental information analysis. *Journal of Accounting research*, *31*(2), 190-215.

- MARTIKAINEN, T. (1989). Modelling stock price behaviour by financial ratios. *Rivista di matematica per le scienze economiche e sociali*, *12*(1), 119-138.

- OMRAN, M., & RAGAB, A. (2004). Linear versus non-linear relationships between financial ratios and stock returns: Empirical evidence from Egyptian firms. *Review of Accounting and finance*, *3*(2), 84-102.

- PATİN, J. C., RAHMAN, M., & MUSTAFA, M. (2020). Impact of total asset turnover ratios on equity returns: Dynamic panel data analyses. *Journal of Accounting, Business and Management (JABM)*, *27*(1), 19-29.

- PESARAN, M.H., (2004), General Diagnostic Test for Cross Section Dependence in Panels, Forschungsinstitut Zur Zukunft Der Arbeit Institute for The Study of Labor, https://docs.iza.org/dp1240.pdf, [Date Accessed: 08.05.2024].

- SHRESTHA, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, *8*(2), 39-42.

- SENCHACK, A.J. & MARTIN, J.D., (1987), Relative Performance of the PSR and PER Investment Strategies, *Financial Analysist Journal,* 43(2), 46-56.

- TATOĞLU, F.Y., (2016), Panel Veri Ekonometrisi, Beta Basım Yayım Dağıtım, İstanbul.

- TEKİN, B., & BASTAK, S. N. (2022). Panel Kantil Regresyon Yaklaşımı İle Getiriyi Etkileyen İçsel Faktörlerin Modellenmesi: BIST 100 Örneği. *Karamanoğlu Mehmetbey Üniversitesi Sosyal ve Ekonomik Araştırmalar Dergisi*, *24*(42), 194-208.

- TOPALOĞLU, E. E. (2018), Finansal Riskler İle Firma Değeri Arasındaki İlişkinin Belirlenmesi: Borsa İstanbul Firmaları Üzerine Bir Uygulama. *Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *5*(2), 287-301.

- TUAN, K., & BORAK, M. Kazanç Yönetimi ve Finansal Performans: Borsa İstanbul Üzerine Ampirik Bir Araştırma. *Sosyoekonomi*, *28*(44), 381-393.

- UTAMI, W. B. (2017). Analysis of Current Ratio Changes Effect, Asset Ratio Debt, Total Asset Turnover, Return On Asset, And Price Earning Ratio In Predictinggrowth Income By Considering Corporate Size In The Company Joined In LQ45 Index Year 2013-2016. *International Journal of Economics, Business and Accounting Research (IJEBAR)*, *1*(01).

- UYAR, U., & SARAK, G. (2020). Finansal Oranlar ile Firma Değeri İlişkisinin Borsa İstanbul ve Londra Borsası İmalat Sanayi Sektörlerinde Karşılaştırılması. *Ekonomi Politika ve Finans Araştırmaları Dergisi*, *5*(3), 537-560.

- ÜNLÜ, U., BAYRAKDAROĞLU, A., & EGE, İ. (2009). Hisse senedi endeks getirileri ve temettü verimi: İMKB 100 ve S&P 500 endeksleri üzerine bir uygulama. *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *11*(1), 143-158.

"This page is left blank for typesetting"

HOLISTENCE publications

JAME
E-ISSN:2791-7401
Volume : 4, Issue 1, 2024
URL: https://journals.gen.tr/jame

Received: 2024-05-09
Accepted: 2024-05-27

RESEARCH ARTICLE

Volatility spillovers of global economic policy uncertainty and fear index among cryptocurrencies: A wavelet-based DCC-GARCH approach

Aslan Aydoğdu

Lec. Dr., Sivas Science and Technology University, Projects Office General Coordinator, Türkiye, email: aaydogdu@sivas.edu.tr

Abstract

This research analyzes the dynamic relationships between the economic and political uncertainty index and the fear index in global markets and cryptocurrencies using the wavelet-based DCC-GARCH method, considering different time scales. Monthly data sets for the periods 2012–April 2024 for GEPU,VIX, and Bitcoin and April 2016–April 2024 for Ethereum are used in the study. Findings are obtained in terms of the volatility interaction between cryptocurrencies (Bitcoin and Ethereum) and GEPU and VIX, as well as four different time scales representing the short, medium, and long term. As a result of the analysis based on raw data, it was found that there is no volatility interaction between cryptocurrencies and GEPU and VIX returns. However, there is a volatility interaction between past volatility shocks and current period volatility shocks in the 4-8 and 16-32 month investment cycle periods of VIX, Bitcoin, GEPU, and Ethereum and time scales. These results, which show that volatility shocks persist in both 4-month and 16-month investment cycles, have significant implications for investors and policymakers. They highlight the need for comprehensive information about changes in the global economy and politics, and they are expected to provide insights for both investors and policymakers.

Keywords: Wavelet Analysis, DCC-GARCH, Volatility Spillovers, Cryptocurrencies, Bitcoin

JEL codes: C10, G11, C53

Citation: AYDOĞDU, A. (2024). Volatility spillovers of global economic policy uncertainty and fear index among cryptocurrencies: A wavelet-based DCC-GARCH approach. *Journal of Applied Microeconometrics (JAME)*. 4(1),13-29, DOI: 10.53753/jame.2417

Corresponding Author:
Aslan Aydoğdu
E-mail: aaydogdu@sivas.edu.tr

1. INTRODUCTION

The concept of investment is generally divided into capital investments and financial investments, and both methods aim to transform savings into capital accumulation. In this context, today's investors prefer many alternative investment instruments to obtain more capital accumulation. However, recently, there has been an orientation towards a different investment instrument. One of these alternative investment instruments is cryptocurrencies. The reasons for the orientation towards cryptocurrencies are technological developments in the payments system, which is a necessity of the digital age that awaits us in the future, and the changing risk perception (Tuncel and Gürsoy, 2020:2000).

The perception that the dollar, which was seen as the safest currency in global markets until the 2008 mortgage crisis, would be dethroned with the crisis, and Nakamoto, who took advantage of this gap, proposed Bitcoin in 2008 and in the following years, different cryptocurrencies such as Ethereum and Ripple emerged. In addition to these cryptocurrencies, the number of cryptocurrencies in actual circulation in the cryptocurrency market is more than 22,000 as of March 2023, with a market capitalization exceeding 1 trillion dollars, showing a significant growth potential (CoinMarketCap, 2023; Aydoğdu, 2024: p.1). Cryptocurrencies, whose functions as money were initially discussed, have recently started to be evaluated as an investment instrument and have been accepted as an alternative investment instrument by a significant mass of investors. With the impact of recent developments, such as; the pandemic and, the Russia-Ukraine crisis, the relationship between the crypto asset market and conventional financial assets has become curious. Especially after the Russia-Ukraine crisis, many people worldwide have questioned financial markets, and decentralized currencies have become a more popular alternative.

Volatility of financial instruments is one of the most important indicators that investors pay attention to when making investment decisions. In addition to the volatility in the national market, investors also follow the volatility in the international market. With the acceleration of globalisation, financial markets are interrelated with each other and a volatility occurring in one of them affects the others. Therefore, investors also take into account the volatility in international markets when making decisions. In this context, the VIX fear index is a volatility index considered by investors (Akdağ, 2019: 236).This index is a risk indicator calculated by the Chicago Options Exchange in the United States based on the differences between the option bid and option ask prices of stocks.

The Volatility Index is an indicator that measures anxiety and fear in the markets and is also known as the VIX Fear Index. Various forecasting methods have been developed to assess the uncertainties in the global economy, and new indexing methods have been emphasised in academic studies. In particular, index that analyse economic and political uncertainties stand out as methods that examine political situations as well as financial risks. The Global Economic Political Uncertainty Index (GEPU), which is the basis of this study, is an index developed by Baker et al. in 2013, representing the economies of 16 countries. This index includes national EPU index based on the frequency of newspaper reports on the economy, uncertainty and politics, and currently includes indices for 21 countries. These 21 countries represent approximately 71 per cent of global output adjusted for purchasing power parity and 80 per cent at market exchange rates (Keser, 2022: p. 121).

Investors portfolio diversification strategies to manage financial risk may vary depending on their investment horizons (short, medium, and long-term). Accordingly, active investors, such as risk-seeker institutional and retail investors, are interested in the interactions between high-frequency (low time-scale) cryptocurrencies and the returns of economic and political uncertainty and fear indices, i.e., short-term fluctuations. On the other hand, passive investors, such as risk-averse and risk-neutral institutional and retail investors, are interested in the interactions between cryptocurrencies at low frequencies (high time-scale) and the returns of the economic political uncertainty index and the fear index, i.e., long-term fluctuations. As a result, investors from different groups face different risks. Studying the relationships between cryptocurrencies and the returns on the economic, political uncertainty, and fear indexes at different time scales is crucial for risk management. As a matter of fact, for investors seeking alternative investment instruments, determining the time scale in which the correlation between the financial assets in the portfolio is low will ensure that the investor will benefit from portfolio diversification (Benhmad, 2013) because a high correlation between assets in a portfolio may cause the investor to make meager gains in terms of risk management. The heterogeneity caused by investors with different investment horizons in the market may cause the spillover effects between markets to change over time and at different frequencies. However, studies examining the relationship between cryptocurrencies and the economic political uncertainty index and the fear index mainly focus on a single time scale and ignore the risks that investors may face according to different time scales and the spillovers between these risks (Uyar and Kangalı Uyar, 2021: p. 310)

With wavelet decomposition analysis, it is possible to examine the change in the relationship between two-time series according to time and different frequencies. Thus, the wavelet approach can help to reveal potential spillover effects by allowing the existence of spillover effects between cryptocurrencies and the economic political uncertainty index, and the fear index returns to be examined according to different time scales. There are many studies on cryptocurrencies in the literature. However, studies on the volatility spillovers between cryptocurrencies and the global economic uncertainty and fear indexes are limited. In addition, no study has been found in the literature that analyses the interaction between cryptocurrencies and GEPU and VIX returns according to different time scales. In short, such a study on cryptocurrencies, whose popularity is increasing day by day and which are at the centre of various debates, will be beneficial to the literature. In addition, analyses based on the combination of both methods can provide inferences on determining the appropriate time periods to benefit from the advantages of portfolio diversification.

Therefore, the purpose of this study is to examine the impact of price movements in the Global Economic Political Uncertainty Index (GEPU) and the Fear Index (VIX) on crypto currencies. For this purpose, in the first section, information on cryptocurrencies, the economic policy uncertainty index (EPI), and the fear index (VIX) are given. A literature review is included in the second part of the study, and studies on volatility spillovers are included. The third section explains information about the data used in the analysis, wavelet decomposition analysis, and the DCC-GARCH model theory. In the fourth section, the findings obtained are interpreted. In the last section, some evaluations are made as a result of these findings.

2. LITERATURE REVIEW

Studies have carried out a comprehensive examination of the relationship between cryptocurrencies and other economic indicators in different time periods and using various analysis methods. These studies aimed to analyse the effects of economic indicators on cryptocurrency markets in detail with the variety of data sets used. However, all of these methods are based on a single time scale. However, recently, studies using methods that take into account different time scales and/or different investment horizons have also taken their place in the literature. The literature on the subject is presented in Table 1.

Table 1. Literature Review

Imprint	Aim	Sampling Period	Method/ Model	Result
Corbet vd. (2018)	It aims to investigate the relationships and frequency fields between crypto currencies and different financial assets.	Daily data set for the periods 29.04.201307.02.2014 10.02.201430.04.2017.	Diebold and Yilmaz analysis method/Barunik and Krehlik method of analysis	It is concluded that cryptocurrencies can provide diversification benefits for investors with short investment horizons. Moreover, time variation in linkages is found to reflect external economic and financial shocks.
İltaş (2020)	This research aims to examine the relationship between BIST 100 Index and economic, political, financial and geopolitical risks.	Monthly data set for the period January 1999-December 2014.	Toda-Yamamoto causality test/ Hacker Hatemi-J bootstrap causality test	It has been concluded that economic, political and geopolitical risks have a symmetric and asymmetric causal relationship with stock prices in Turkey.

Bakır (2021)	This research aims to examine the relationship between cryptocurrencies and economic indicators.	Daily data set for the period 03.12.2019-03.20.2020.	Pedroni cointegration test Granger causality test	As a result of the analysis, it is concluded that Bitcoin and Ethereum do not have a long-run relationship with some commodities and economic indicators. However, bidirectional and unidirectional causality relationships were found between Bitcoin and Ethereum and G20 stock market indices, some commodities and volatility indices. Moreover, these cryptocurrencies are found to have significant regression relationships with market volumes, some commodities and economic indicators.
Khan vd. (2021)	It aims to examine the relationship between global economic policy uncertainty and bitcoin prices.	Monthly data set between April 2011 and March 2020.	Rolling window method Granger causality	According to the results of the analyses, it is determined that there is no causality relationship between GEPU and BCP. However, considering the structural changes, it is concluded that the full sample causality relationship between the variables may be different. Moreover, the finding of the rolling window test shows that there is causality in different sub-samples; both positive and negative bidirectional causality between GEPU and BCP were found in these sub-samples.
Gürsoy ve Kılıç (2021)	This research aims to analyse the impact of global economic and political uncertainties on financial markets in Turkey.	Monthly data set for the period March 2010-October 2020.	DCC-GARCH	The analyses reveal that there is a strong volatility relationship between GEPU index, CDS premium and BIST banking index.
Sajeev and Afjal (2021)	It is aimed to examine the contagion effect of Bitcoin on the National Stock Exchange, Shanghai Stock Exchange, London Stock Exchange and Dow Jones Industrial Average by analysing the volatility spread and correlation between these markets.	Daily data set for the period March 2017-May 2021.	BEKK-GARCH / DCC-GARCH	The overall time-varying correlation between Bitcoin and stock markets is low, indicating that Bitcoin can be taken as an asset to hedge against the risk of these stock markets. It was also concluded that these stock markets reacted more to negative shocks than to positive shocks in the Bitcoin market in 2018 and 2021.
İmre (2021)	It is aimed to examine the volatility interaction between Bitcoin and Euro returns.	Daily data set for the period 02 February 2014-28 February 2021	BEKK-GARCH / DCC-GARCH	The analyses revealed a bidirectional volatility interaction between the Euro and Bitcoin. In addition, an asymmetry relationship and a positive, strong dynamic correlation between the two returns were found.
Ghorbel and Jeribi (2021)	It aims to analyse the relationships between five cryptocurrencies and the volatilities of S&P500, Nasdaq and VIX indices, oil and gold.	Daily data set for the period 01 January 2016 - 01 April 2020	BEKK-GARCH / DCC-GARCH	The results of both analyses show evidence of a higher volatility spread between cryptocurrencies and a lower volatility spread between cryptocurrencies and financial assets, and the introduction of Bitcoin futures is found to have a significant impact.

Gökalp (2022)	This research aims to investigate the impact of cryptocurrency market developments on Borsa Istanbul (BIST) indices.	Daily data set for the period 01/01/2014-31/12/2021.	BEKK-GARHCH/DCC-GARCH	According to the analysis results, a positive spillover effect from the cryptocurrency markets to indices has been identified. Oil prices, as one of the control variables, have shown a significant impact on volatility across all models. Furthermore, varying results were observed concerning the influence of the fear index.
Keser (2023)	This research aims to investigate the causality relationship between global economic political uncertainty and geopolitical risk and Bitcoin energy consumption.	Monthly data set from May 2011 to February 2022.	Lee-Strazich unit root test Hatemi-J (2012) causality test	According to the analysis results, it has been determined that global economic political uncertainty and geopolitical risk have an impact on bitcoin energy consumption. It has also been concluded that the negative effects of global uncertainty and geopolitical risks are more dominant.

3. METHODOLOGY

3.1. Purpose of the Research and Data Set

The main objective of this research is to examine the impact of global economic and political uncertainties and changes in the fear index on cryptocurrencies. In this context, the global economic and political uncertainty index (GEPU) developed by Baker et al. (2016) and Davis (2016), the fear index (VIX) created by CBOE, and the cryptocurrencies Bitcoin and Ethereum are determined as the main variables used in this study. Data for the variables were obtained by using the "policyuncertainty.com" database for the global economic uncertainty index. Fear index, Bitcoin and Ethereum data were obtained using the "investing.com" database. In the study, the period between GEPU, VIX and Bitcoin is April 2012-April 2024; For Ethereum, the wavelet-based DCC-GARCH model was run using monthly data between April 2016 and April 2024. While creating the data set of the study, these dates were determined due to data constraints for the variables. In the study, the Wavelet-based DCC-GARCH model determines the volatility interaction and transfer according to different time scales and also indicates the relationship between variables. For this reason, the DCC-GARCH model based on Wavelet was preferred in the study. The returns of the variables were calculated according to the formula in equation (1):

$$r_{i,t} = \log\left(\frac{P_{i,t}}{P_{i,t-1}}\right) * 100 \quad (1)$$

Here $r_{i,t}$, represents the return series of the i'th financial asset at time t; $P_{i,t}$ and $P_{i,t-1}$ show the closing prices of financial assets at time t and t-1, respectively.

3.2. Wavelet Decomposition Analysis

Wavelets can be explained as short fluctuations in time with a specific start and end point (Lehkonen and Heimonen, 2014: p.92).Although the Wavelet approach is more comprehensive than the Fourier approach, it is an approach that allows the behavior of a time series to be separated and examined according to different frequencies over time. This flexibility provided by the wavelet approach allows revealing time series behavior or features that cannot be revealed with only time-dependent approaches. This is because it can be examined how the relationships between variables change over time according to different frequencies. Thus, instead of examining financial asset returns for different periods in finance applications, various return layers that make up the total return can be examined. Similarly, instead of examining volatility for different periods, various volatility layers that make up volatility can be obtained through wavelet decomposition and how the behavior of risk evolves over time can be observed. The wavelet approach allows time series to be analyzed without applying any transformation to non-stationary time series and therefore without loss of observation (Schleicher, 2002:p.27).

Moreover, the non-parametric nature of the wavelet approach allows non-linear relationships between variables to be taken into account without any loss of information. These advantages of the wavelet approach, which is based on time and frequency, are the most important reasons why it is a more effective technique than time-only approaches. The wavelet approach was first applied in the field of economics by Ramsey and Lampart (1998a,b) in order to analyze the relationships with money supply (M1 and M2). In recent years, Rua and Nunes (2009), Jammazi and Aloui (2010), Masih etal., (2010), Ismail et al., (2016), Omane-Adjepong and Alagidede (2019), Uyar and Kangallı Uyar (2021), Hairudin and Mohamad (2023), Aydoğdu (2024) etc. It has been introduced to the literature in economics and finance by researchers such as using this approach.

In wavelet analysis, a time series can be decomposed into different time scales by applying wavelet transformation. Father wavelet in wavelet analysis Two basic functions are defined: mother wavelet and mother wavelet. The father wavelet contains the low-frequency components of the original series and shows the trend of the series; The mother wavelet contains the high-frequency components of the series and shows deviations from the trend, in other words, it reflects the details in the data (Crowley, 2007). The father wavelet and mother wavelet can be defined as in equation (2) and equation (3), respectively (Ramsey and Lampart, 1998):

$$Ø_{j,k}(t) = 2^{-\frac{1}{2}} Ø\left(2^{-j} * t - k\right), j = 1,2, \dots, J; k = 0,1, \dots, 2^j - 1 \quad (2)$$

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}} \psi\left(2^{-j} * t - k\right), j = 1,2, \dots., J; k = 1, \dots., 2^j - 1 \quad (3)$$

Wavelet functions depend on the scale or frequency parameter j and the location parameter k. Although 2^j, is called the scale, it can be treated as a measure of the width of the $Ø_{j,k}$(t) function. Accordingly, as the values of j increase, the function becomes shorter and spreads further. The scale parameter determines the size of the wavelet, while the position parameter sets the location of the wavelet. A scale parameter ranging from 1 to J means that the time series is decomposed at J different levels according to the highest time scale J. $Ø(.)$ ve $\psi(.)$, (-∞,∞) are real-valued functions defined on the real axis, and it is assumed that these functions meet the normalization conditions defined in equations (4) and (5):

$$\int_{-\infty}^{+\infty} Ø(t)dt = 1 \quad (4)$$

$$\int_{-\infty}^{+\infty} \psi(t)dt = 1 \quad (5)$$

A time series such as x(t) defined at $L^2(R)^3$ can be expressed in terms of wavelet functions as in equation (6):

$$x(t) = \sum_{k=0}^{2^j-1} S_{J,k} Ø_{J,k}(t) + \sum_{j=1}^{J} \sum_{k=0}^{k^j-1} d_{j,k}\ \psi_{j,k}(t) \quad (6)$$

Here it is defined as$S_{J,k} = \int_{-\infty}^{+\infty} t(t) Ø_{J,k}(t)dt = 1$ ve $d_{j,k} = \int_{-\infty}^{+\infty} x(t)\psi_{j,k}(t)dt = 1$. $S_{J,k}$, are called smooth coefficients, while $d_{j,k}$ are called detail functions. The sizes of these coefficients show the share of wavelet functions in the total data. In the expression in Equation (7);

$$S_t(t) = \sum_{k=0}^{2^j-1} S_{J,k} Ø_{J,k}(t) \quad (7)$$

ve

$$D_j(t) = \sum_{k=0}^{2^j-1} d_{j,k}\, \psi_{j,k}(t), j = 1,2, \dots.., J \quad (8)$$

When defined as x(t) time series can be re-expressed as in equation (9):

$$x(t) = S_J(t) + \sum_{j=1}^{J} D_j(t) \quad (9)$$

Here, $S_J(t)$, reflects the trend of the data; because it is the component of the highest level time scale. $D_j(t) = (D_1(t), D_2(t), \dots \dots \dots \dots \dots \dots, D_J(t))$ are the details containing the fluctuations in the data on 2-4, 4-8,…..,$2^J - 2^{J+1}$ time scales, respectively.

Small values of j correspond to the low time scale, thus representing the high-frequency components of $x(t)$, while large values of j correspond to the high time scale, and thus represent the low-frequency components of $x(t)$. Since D_j, includes cyclical movements between coefficients $2^J - 2^{J+1}$,in period D_1, 2-4; In period D_2, 4-8, etc. Includes cyclic movements.

Discrete wavelet transformation can be applied to obtain wavelet coefficients (smooth and detailed coefficients). In discrete wavelet transform (DWT), the researcher determines how many different time scales the time series will be decomposed according to the number of observations. Moreover, in DWT, it is stated that the number of observations must have a dyadic feature, in other words, it must be an integer that is a multiple of two. Since this feature of DWT is restrictive in determining time scales in applications, *"maximum overlap discrete wavelet transform (MODWT)",* a type of discrete wavelet transform, is generally applied instead of DWT. On the other hand, any sample size can be used for MODWT, it does not have to be dyadic. In this study, MODWT will be used to obtain wavelet coefficients due to the restrictive assumptions of DWT.

Wavelet filters with different features (D(4), D(8), LA(8) and Haar wavelet) can be used to obtain wavelet coefficients with MODWT (Gençay et al., 2002: 113-116). Gençay et al. (2002, 2010) and Cornish et al. (2006) and others and frequently used in the literature is LA(8) (least asymmetric wavelet filter of length eight). This is because the LA(8) filter produces smoother and uncorrelated wavelet coefficients than other filters.

In the study, after the wavelet coefficients based on MODWT were calculated for the returns of each variable, the return series were calculated according to four different time scales (D_1-2-4 months, D_2-4-8 months, D_3-8-16 months, Dg=16-32 months, Sg) was isolated. Cycle times for different time scales are defined in Table 2:

Table 2. Wavelet Analysis Time Horizons According to the Multiple Scale Method

Scales (2^j)		Annual Frequency	Monthly Frequency	Daily Frequency
1	2^1	2-4	2-4	2-4
2	2^2	4-8	4-8	4-8
3	2^3	8-16	8-16 (8 months-1 year 4 months)	8-16
4	2^4	16-32	16-32 (1 years 4 months-2 years 8 months)	16-32 (3 weeks 1 day-6 weeks 2 day)
5	2^5	32-64	32-64 (2 years 8 months-5 years 4 months)	32-64 (6 weeks 2 day-12 weeks 4 day)
6	2^6	64-128	64-128 (5 years 4 months-10 years 8 months)	64-128 (12 weeks 4 day-25 weeks 3 day)
7	2^7	128-256	128-256 (10 years 8 months-21 years 4 months)	128-256 (25 weeks 3 day- 51 weeks 1 day)
8	2^8	256-512	…..	….

Source: Crowly (2007:214). Theoretically, the maximum number of scales is expressed as 9. In the case where the scale number is indicated by j (j = 9), the frequencies are calculated using 2j notation.

Considering the scale frequencies given in Table 2; After a wavelet analysis is performed, predictions will be made for the specified number of scales and coefficients will be estimated for various investment horizons. The time scales in Table 2 are grouped as follows: $(D_1 - D_2 - D_3)$: short term; $(D_4 - D_5 - D_6)$: medium term; $(D_7 - D_8 - S_8)$: long term. This type of grouping was made to examine how the movements of investors with short, medium and long-term investment horizons develop according to different time scales. Short-term investment horizons $(D_1 - D_2 - D_3)$; It refers to short-term changes due to shocks occurring on time scales of 2-16 months and includes daily-weekly spreads. Medium-term investment horizons $(D_4 - D_5 - D_6)$; It shows medium-term changes on time scales of 32-128 months and covers monthly to quarterly spreads. Long-term investment horizons are $(D_7 - D_8 - S_8)$; It indicates long-term changes on time scales of 256 months and longer and is a period covering annual spreads (Uyar and Kangalı Uyar, 2021: p. 319; Aydoğdu, 2024: p. 217-218).

3.3. DCC-GARCH Approach

The dynamic conditional correlation (DCC) approach was developed by Engle (2002) to examine time-varying correlations between asset returns. To examine in more detail how this approach was developed and what assumptions it is based on, let the vector containing the logarithmic returns of k financial assets be denoted by r_t. Financial asset returns have the following distribution under the assumptions that there is no autocorrelation between average returns and that quadratic moments vary over time:

$$r_t \mid I_{t-1} \sim D(\mu, H_t) \tag{10}$$

Here, r_t, is the payoff vector of size I_{t-1},; $t-1$ represents the information set at time t-1; μ is the unconditional mean, which is usually very close to or equal to zero; H_t. denotes the dynamic conditional covariance matrix of the k-return series of size kxk, and $D(u, H_t)$, denotes the multivariate density function that depends on the mean vector and the dynamic conditional covariance matrix.

Engle (2002; p. 341) decomposed the covariance matrix as the product of dynamic conditional standard deviations and dynamic conditional correlations:

$$H_t = D_t R_t D_t \tag{11}$$

Here, D_t, is a diagonal matrix of size kxk and its elements consist of time-varying standard deviations obtained from univariate GARCH models. In the D_t matrix, the i'th element of the diagonal can be represented as $\sqrt{h_{it}}$ hit:

$$D_t = \begin{bmatrix} \sqrt{h_{1t}} & 0 & \cdots & 0 \\ 0 & \sqrt{h_{2t}} & \ddots & \vdots \\ 0 & \dots & \cdots & \sqrt{h_{kt}} \end{bmatrix} \tag{12}$$

R_t, is the time-varying correlation matrix of $\eta_t = D_t^{-1} * r_t, R_t \sim N(0, R_t)$, standardized residues of size kxk:

$$R_t = \begin{bmatrix} 1 & \rho_{12,t} & \dots & \cdots & \rho_{1k,t} \\ \rho_{12,t} & 1 & \vdots & \ddots & \vdots \\ \rho_{1k,t} & \dots & \rho_{k-1,k,t} & \cdots & 1 \end{bmatrix} \tag{13}$$

Since H_t, is a covariance matrix, it must be a positive definite matrix. Since D_t, is a positive definite matrix due to its positive diagonal elements R_t must also be a positive definite matrix. Finally, the elements in the must be less than or equal to 1 because they include conditional correlation coefficients.

Using the representations in Equations (10)-(11), it can be deduced that the marginal density function of each element of the r_t vector, in other words, the return of each financial asset, *"depends on the time-varying conditional variance"* and the time-varying conditional variance, which is the representative of volatility. can be modeled as a univariate GARCH process. Accordingly the D_t matrix can be created using the univariate GARCH(p, q) model defined in equation (11):

$$h_{it} = \theta_i + \sum_{p=1}^{p_i} \alpha_{ip}\, \varepsilon_{it-p}{}^2 + \sum_{q=1}^{Q_i} \beta_{iq} h_{it-q} \quad i = 1,2, \ldots\ldots, k \tag{14}$$

Here, θ_i is the constant term. non-negativity of parameters and stationarity in variance in the GARCH(p, q) model:

$$\sum_{p=1}^{p_i} \alpha_{ip}\, \varepsilon_{it-p}{}^2 + \sum_{q=1}^{Q_i} \beta_{iq} h_{it-q} < 1 \tag{15}$$

It is assumed that the constraints in equality (15) are satisfied. Satisfying these constraints ensures that the matrix is a positive definite H_t matrix for all time points.

The standardized return vector is obtained by dividing the return of each financial asset by its conditional standard deviation, $\sqrt{h_{it}}$,:

$$\eta_t = D_t{}^{-1} * r_t, \qquad \eta_t \sim N(0, R_t) \tag{16}$$

This vector can be used to define the dynamic conditional correlation structure defined by Engle (2002) for the return series:

$$Q_t = \left(1 - \sum_{m=1}^{M} \alpha_m - \sum_{n=1}^{N} \beta_n\right) * \bar{Q} + \sum_{m=1}^{M} \alpha_m (\eta_{t-m} \acute{\eta}_{t-m}) + \sum_{n=1}^{N} \beta_n Q_{t-n} \tag{17}$$

Here α_m and β_n are non-negative scalars. Due to the stability condition, it is assumed that $\alpha_m + \beta_n < 1$. Providing these assumptions is important for the H_t matrix to be a positive definite matrix. Since $Q_t = \bar{Q}$ if the constraint $\alpha_m = \beta_n = 0$ is met, it can be said that the use of a time-invariant conditional correlation model will be sufficient to examine the relationships (Lebo and Box Steffensmeier, 2008: 694). $\bar{Q}$, is the unconditional covariance matrix of the standardized residuals from the first-stage estimation:

$$\bar{Q} = Cov[\eta_t \acute{\eta}_t] = E[\eta_t \acute{\eta}_t] \text{ ve } \bar{Q} = \tfrac{1}{T} \textstyle\sum_{t=1}^{T} \eta_t \acute{\eta}_t \tag{18}$$

It can be predicted by. Finally, the model defined in equation (17) can be represented as DCC(m, n).

However, since Q_t, does not meet the definition of a dynamic conditional correlation matrix, Engle (2002) suggested the following standardization:

$$R_t = Q_t{}^{*-1} * Q_t * Q_t{}^{*-1} \tag{19}$$

Here $Q_t{}^*$, is a diagonal matrix containing the square roots of the diagonal elements of the matrix $\bar{Q}_t$:

$$Q_t{}^* = \begin{bmatrix} \sqrt{q_{11,t}} & 0 & \cdots & 0 \\ 0 & \sqrt{q_{22,t}} & \ddots & \vdots \\ 0 & \cdots & \cdots & \sqrt{q_{nn,t}} \end{bmatrix} \tag{20}$$

$Q_t{}^*$, rescales the elements of the Q_t matrix to be $|\rho_{ijt}| = \left|\frac{q_{ijt}}{\sqrt{q_{iit} q_{jjt}}}\right| \leq 1$. Q_t must be a positive definite matrix.

The elements of the R_t matrix are in the form $\rho_{ijt} = \frac{q_{ijt}}{\sqrt{q_{iit} q_{jjt}}}$. The fact that the R_t matrix is a positive definite matrix depends on the Q_t matrix being a positive definite matrix.

Engle (2002) defined the logarithmic likelihood function for the maximum likelihood estimation of the DCC model defined in equation (12) as follows:

$$L = \frac{1}{2}\sum_{t=1}^{T} k * \log(2\pi) + \log(|H_t|) + r_t' H_t^{-1} r_t \quad (21)$$

If the expression where H_t is equal in equation (10) is substituted in equation (15), the logarithmic likelihood function in equation (16) is obtained by using $\eta_t = D_t^{-1} * r_t$ equation:

$$L = \frac{1}{2}\sum_{t=1}^{T} k * \log(2\pi) + 2 * \log(|D_t|) + n_t' R_t^{-1} n_t \quad (22)$$

The definition of the logarithmic likelihood function in equation (22) facilitates the estimation of the DCC model, because the function has two components, namely the volatility component and the correlation component, and the estimation can be carried out by splitting the estimation process into two. The first component is the volatility component and contains only terms in D_t, and the second component is the correlation component and contains only terms in R_t, The reason why the DCC model is estimated in two stages can be explained by this structure. In the first stage, only the part containing the volatility component is maximized, and in the second stage, the correlation component conditional on D_t is maximized, and thus estimates of the parameters of the DCC model, α_m and β_n are obtained. The parameters α_m and are the determinants of the correlation between two series. The parameter shows the short-term effects of volatility, and the β_n parameter shows the long-term permanent effects (Uyar and Kangalı Uyar, 2021: p. 322) .

4. ANALYSIS FINDINGS

In this study using monthly data, descriptive statistics and unit root analysis of the raw returns of the variables were performed. Then, the study analyzes were carried out in two stages. In the first stage, variable returns were decomposed into different time scales using wavelet decomposition analysis. In the second stage, the dynamic correlation relationship between the return series of variables separated according to different time scales was examined according to the DCC-GARCH approach. The DCC-GARCH (1,1) model to be estimated was created as follows:

$$r_t = \mu + \varepsilon_t, t = 1,2,\dots,T, \varepsilon_t \mid I_{t-1} \sim N(0, H_t) \quad (23)$$

$$H_t = D_t R_t D_t, D_t = diag\left(\sqrt{h_{1t}}, \sqrt{h_{2t}}\right) \quad (24)$$

$$h_{it} = \theta_i + \alpha_i \varepsilon_{it-1}^{2} + \beta_i h_{it-1}, i = \text{VIX, Bitcoin} \quad (25)$$

$$R_t = Q_t^{*-1} * Q_t * Q_t^{*-1}, Q_t^{*} = diag\left(\sqrt{q_{11,t}}, \sqrt{q_{22,t}}\right) \quad (26)$$

$$Q_t = (1 - \alpha - \beta) * \bar{Q} + \alpha(\eta_{t-1}\acute{\eta}_{t-1}) + \beta Q_{t-1} \quad (27)$$

Here, the DCC-GARCH (1,1) model is defined to study the dynamic correlation between VIX and Bitcoin. Similarly, the DCC-GARCH (1,1) model is created for other variable returns. DCC-GARCH (1,1) model is defined within the return series of the variables separated into different time scales. These models are defined as wavelet-based DCC-GARCH(1,1) models. In the models, VIX return on time scale D_1 and Bitcoin on time scale D_1, VIX return on time scale D_4, and Bitcoin on time scale , etc. The dynamic correlations between them were examined. Similar matchings were made for return series of variables on other time scales. As a result, 4 DCC-GARCH (1,1) models based on raw data; 16 DCC-GARCH (1,1) models based on four different time scales (4x4) were estimated.

Table 3. Descriptive Statistics of Return Series of Variables

	Bitcoin	Ethereum	GEPU	VIX
Mean	0.0651	0.0604	0.0026	0.0004
Median	0.0277	0.0316	-0.0020	-0.0040
Maksimum	1.7249	1.1426	0.6279	0.8525
Minimum	-0.4665	-0.7719	-0.4987	-0.6142
Standard Deviation	0.2656	0.3135	0.1791	0.2421
Skewness	1.8256	0.4579	0.4104	0.3971
Kurtosis	12.9327	4.0861	4.5647	3.8066
Jarque-Bera	681.2857***	8.1582**	4.6547***	7.7959**
P. Value	[0.0000]	[0.0169]	[0.0000]	[0.0202]
Observations	146	97	146	146

Note: ***, ** and * indicate statistical significance at 1%, 5% and 10% confidence intervals, respectively.

Table 3. includes descriptive statistics of the return series of Bitcoin, Ethereum, Global Economic Political Uncertainty Index (GEPU) and Fear Index (VIX). It can be seen that the mean values of all variables are close to zero and positive. While the highest average return belongs to Bitcoin and Ethereum; When evaluated in terms of volatility, it can be seen that Bitcoin has the highest standard deviation in the examined data range and is the GEPU with the lowest standard deviation. It can be seen that the coefficients of all return series are positive and right-skewed. It is seen that the kurtosis values are positive and greater than three and the series have an extremely flat (leptokortic) structure. Positive skewness coefficients indicate that positive returns occur more frequently than negative extreme returns. When the Jarque-Bera test statistics, which test the normality assumption of all variables, were evaluated, it was confirmed that the return series did not comply with the normal distribution. This result provides an important reason for using the wavelet-based DCC-GARCH approach, which does not make any distributional assumptions, to examine the relationships between markets. Before performing the wavelet-based DCC-GARCH analysis, unit root tests were carried out to determine whether the return series were stationary.

Table 4. Unit Root Tests of Price and Return Series of Variables

		Price			**Return**		
	Model	ADF	PP	KPSS	ADF	PP	KPSS
Bitcoin	Constant	-0.7065	-0.1610	1.1851	-9.5077***	-9.4382***	0.1863
	Constant and Trend	-2.5251	-1.9965	0.1686	-8.9934***	-9.6142***	0.0469
Ethereum	Constant	0.7220	-1.1838	0.8735	-8.2236***	-8.5232***	0.0766
	Constant and Trend	-2.1279	-2.2658	0.0869	-8.2061***	-8.4822***	0.0568
GEPU	Constant	-2.9564	-2.6875	0.9488	-15.1740***	-15.9600***	0.2318
	Constant and Trend	-3.1453	-2.8006	0.1558	-15.2393***	-16.0917***	0.1117
VIX	Constant	-5.3098	-5.1646	0.4754	-11.59680***	-27.2617***	0.1622
	Constant and Trend	-5.5897	-5.5029	0.1156	-11.5509***	-27.2626***	0.1593

Note: In the ADF test, the maximum number of delays was taken as 13 and the optimum number of delays was determined according to the Schwarz Information Criterion. Long-term variance in PP and KPSS tests was obtained with the Bartlett kernel estimator and bandwidth was determined with the Newey-West method. In ADF and PP tests, critical values are -3.433122 (1%), -2.862651 (5%) and -2.567407 (10%) for the constant model; for the constant and trend model it is -3.962212 (1%), -3.411849 (5%) and - 3.127817 (10%). In the KPSS test, the critical values for the constant model are 0.739000 (1%), 0.463000 (5%) and 0.347000 (10%); For the constant and trending model, it is 0.216000 (1%), 0.146000 (5%) and 0.119000 (10%). The symbols ***, **, and * indicate statistical significance at 1%, 5% and 10% significance levels.

The long-term characteristics of a time series are revealed by determining how the variable value in the previous period affects the current period. In order to understand the evolution of the time series, it is necessary to perform regression analysis of the values in each period compared to previous periods. The unit root analysis method used to determine the stationarity of the series is an effective tool for evaluating this process (Tarı, 2014: 387; Aydoğdu, 2024: 243). In this context, Table 4 includes Augmented Dickey Fuller (ADF), Phillips-Perron (PP) and

Kwiatkowski, Phillips, Schmidt and Shin (KPSS) unit root test results to determine the stationarity of the series examined within the scope of the research. While ADF and PP tests indicate the case of unit root, non-stationarity or I[1] with "H_0*(null hypothesis)";* KPSS test indicates the I[0] process, in other words, stasis (Şahin Dağlı and Çelik, 2022; 2198). According to the ADF and PP tests applied to Bitcoin, Ethereum, GEPU and VIX returns, it is seen that the unit root H_0 hypothesis is rejected, and for the KPSS test statistic result, the I[0] process is reached and H_0 cannot be rejected. As a result, it was concluded that the series do not contain unit roots and have a stationary structure.

Tables 5, 6, 7 and 8 show the predictions of DCC-GARCH (1,1) models for return series disaggregated according to raw data and four different time scales. The coefficient α shows the effect of standardized shocks $(\eta_{t-1}\acute{\eta}_{t-1})$ and the coefficient β shows the effect of lagged dynamic conditional correlations Q_{t-1} on the dynamic conditional correlations in the current period. Statistically significant values of the α coefficient indicate short-term permanence, while large and statistically significant values of the β coefficient indicate long-term permanence.

Table 5. Analysis Results for VIX-Bitcoin Raw Data and Return Scales

Panel A:					
Raw Data	α	p	β	p	$\alpha+\beta$
Bitcoin	-0.0000	0.9999	0.5972	0.2570	0.5972
Panel B:					
Return Scales					
D1	0.5375	0.0000	0.2277	0.1682	0.7652
D2	**0.2453**	**0.0023**	**0.6544**	**0.0000**	**0.8997**
D3	0.4548	0.0000	-0.0000	0.9999	0.4548
D4	**0.9855**	**0.0000**	**0.0143**	**0.0000**	**0.9998**

Table 5. includes the analysis findings of Fear index (VIX) and Bitcoin raw data and different time scales. According to the findings, it can be concluded that the α and β coefficients of Bitcoin raw data returns are not significant, and that the shocks and dynamic conditional correlations in the past period have no effect on the dynamic conditional correlations of VIX and Bitcoin returns in the current period. This result is also an indication that there is no dynamic relationship or volatility interaction between the Fear index and Bitcoin returns. When the predictions of wavelet-based DCC-GARCH(1,1) models (from to) are examined; It took values of $\alpha+\beta<1$ at all time scales. However, it was concluded that $\alpha+\beta<1$ was statistically significant in the D_2 and D_4 time scales. This finding shows that dynamic correlations fluctuate around a fixed level in 4-8 month and 16-32 month investment cycles and have a process that tends to return to the mean.

The α coefficient estimates of the DCC-GARCH(1,1) models at all time scales are statistically significant at the 1% significance level, but the statistical significance of the β coefficients is significant at the 1% significance level in D_2 and D_4, but the β coefficients are not significant at the D_1 and D_3 time scales. has been observed. When these findings are evaluated, it can be said that past period conditional correlations on D_1 and D_3 time scales do not have an effect on current period correlations, in other words, volatility shocks do not show permanence in 2-4 month and 8-16 month investment cycles. It can even be stated that there is no volatility interaction in these investment cycles. On the other hand, it can be stated that there are time-varying correlations between VIX and Bitcoin returns on the D_2 and D_4 time scales, and that past volatility shocks and conditional correlations are effective on these correlations. In other words, volatility shocks are permanent in both 4-8 month and 16-32 month investment cycles. In addition, in the 4-8 and 16-32 month investment cycles, although $\alpha+\beta<1$, it is very close to 1 (one); with VIX, it can be stated that conditional volatility for Bitcoin is more likely to be permanent.

Table 6. Analysis Results for GEPU-Bitcoin Raw Data and Return Scales

Panel A:					
Raw Data	α	p	β	p	$\alpha+\beta$
Bitcoin	-0.0000	0.9999	0.0891	0.7611	0.0891
Panel B:					
Return Scales					
D1	0.5647	0.0000	0.1465	0.1721	0.7103
D2	**0.4292**	**0.0000**	**0.4567**	**0.0000**	**0.8859**
D3	0.2816	0.0000	-0.0000	0.9999	0.2816
D4	0.2295	0.0000	-0.0000	0.9999	0.2295

Table 6. shows the findings regarding GEPU and Bitcoin raw data and different time scales. According to the findings, the α and β parameter coefficients regarding Bitcoin raw data returns are not statistically significant. Therefore, shocks and dynamic conditional correlations in the past period are an indication that GEPU and Bitcoin returns in the current period have no effect on dynamic conditional correlations. In other words, it can be stated that there is neither a dynamic relationship nor a volatility interaction between the global economic political index and Bitcoin returns. When the predictions of wavelet-based DCC-GARCH(1,1) models (from D_1 to D_4) are examined; The α coefficient estimates of the DCC-GARCH(1,1) models at all time scales are positive and statistically significant at the 1% significance level, but the statistical significance of the β coefficients is only significant at the 1% significance level; However, it was concluded that the β coefficients were not significant at the time scales D_1, D_3 and D_4.

When these findings are evaluated; In the investment cycle periods of 2-4 months, 8-16 months and 16-32 months, past period conditional correlations do not have an effect on current period correlations, in other words, volatility shocks are affected according to the investment cycle periods of 2-4 months, 8-16 and 16-32 months. It can be stated that it is not permanent. It can even be stated that there is no volatility interaction during these investment cycle periods. On the other hand, it can be stated that there are time-varying correlations between GEPU and Bitcoin returns during , that is, the 4-8 month investment cycle period, and that past volatility shocks and conditional correlations are effective on these correlations. In other words, volatility shocks are permanent in the 4-8 month investment cycle. In addition, in the investment cycle period of 4-8 months, although $\alpha+\beta<1$, it is very close to 1 (one); with GEPU, it can be stated that conditional volatility for Bitcoin is more likely to be permanent.

Table 7. Analysis Results for VIX-Ethereum Raw Data and Return Scales

Panel A:					
Raw Data	α	p	β	p	$\alpha+\beta$
Ethereum	-0.0000	0.9999	0.3350	0.1654	0.3350
Panel B:					
Return Scales					
D1	0.3251	0.0469	0.3188	0.4781	0.6439
D2	0.1792	0.2102	0.6596	0.0169	0.8388
D3	-0.0013	0.9589	0.1807	0.0000	0.1794
D4	0.8100	0.0000	-0.0000	0.9999	0.8100

Table 7. includes the Fear index (VIX) and Ethereum raw data and analysis findings for different time scales. According to these findings, the α and β parameter estimation coefficients of Ethereum raw data returns are not statistically significant. According to this result, volatility shocks and dynamic conditional correlations in the past period show that there is no effect on the fear index and Ethereum returns dynamic conditional correlations in the current period. In other words, it is an indication that there is no volatility interaction between the fear index and Ethereum returns. When the estimates of the wavelet-based DCC-GARCH(1,1) models (from D_1 to D_4) were examined, it was determined that the α coefficient estimates of the DCC-GARCH (1,1) models were not statistically significant at all time scales. When the β coefficient estimates are examined, it is not statistically significant in the 2-4 month and 16-32 month investment cycles, but it is statistically significant in the 4-8 month and 8-16 month investment cycles. According to these findings, it has been determined that there is no relationship between the fear index and volatility shocks and dynamic correlations in the past period, volatility shocks and dynamic conditional correlations in the current period in Ethereum returns according to different time scales.

Table 8. Analysis Results for GEPU-Ethereum Raw Data and Return Scales

Panel A:					
Raw Data	α	p	β	p	$\alpha + \beta$
Ethereum	-0.0000	0.9999	0.4132	0.3956	0.4132
Panel B:					
Return Scales					
D1	0.3023	0.0139	0.2152	0.2535	0.5175
D2	**0.3271**	**0.0112**	**0.4845**	**0.0020**	**0.8116**
D3	0.3239	0.0000	-0.0000	0.9999	0.3239
D4	**0.8268**	**0.0000**	**0.0883**	**0.0000**	**0.9151**

Table 8. includes the Global Economic Political Uncertainty Index (GEPU) and Ethereum raw data and analysis findings for different time scales. According to the findings, it can be concluded that the α and β coefficients of Ethereum raw data returns are not significant, and that the shocks and dynamic conditional correlations in the past period have no effect on the global economic political uncertainty index and Ethereum returns dynamic conditional correlations in the current period. This result is also an indication that there is no dynamic relationship or volatility interaction between the global economic political uncertainty index and Ethereum returns. When the predictions of wavelet-based DCC-GARCH(1,1) models (from D_1 to D_4) are examined; It took values of $\alpha+ \beta<1$ at all time scales. However, it was concluded that $\alpha+ \beta<1$ was statistically significant in the D_2 and D_4 time scales. This finding shows that dynamic correlations fluctuate around a fixed level in the 4-8 month and 16-32 month investment cycles and have a process that tends to return to the mean. The α coefficient estimates of the DCC-GARCH(1,1) models at all time scales are positive and statistically significant at the 1% significance level, but the statistical significance of the β coefficients is significant at the 1% significance level in D_2 and D_4, but the β coefficients are significant at D_1 and D_3 time scales. It was observed that it was not significant.

When these findings are evaluated, it can be said that past period conditional correlations have no effect on current period correlations in and time scales, in other words, volatility shocks do not show permanence in 2-4 month and 8-16 month investment cycles. It can even be stated that there is no volatility interaction in these investment cycles. On the other hand, it can be stated that there are time-varying correlations between GEPU and Ethereum returns on the D_2 and D_4, time scales, and past volatility shocks and conditional correlations are effective on these correlations. In other words, volatility shocks are permanent in both 4-8 month and 16-32 month investment cycles. In addition, in the 4-8 and 16-32 month investment cycles, although $\alpha+ \beta<1$, it is very close to 1 (one); with GEPU, it can be said that conditional volatility for Ethereum is more likely to be permanent.

5. CONCLUSION

Cryptocurrencies have recently become an alternative investment tool and an area of regulation that attracts the attention of both individual investors and corporate authorities. However, the high volatility in cryptocurrencies and the impact of some international developments on cryptocurrencies worry investors. In this context, the relationship between crypto currencies and developments involving global economic and political uncertainty has become a matter of curiosity. Whether there is a dynamic interaction between crypto currencies and the global economic political uncertainty index and fear index, and the direction of the relationship, is a phenomenon that should be carefully evaluated by both investors and policy makers. Therefore, the purpose of this study is to examine the impact of price movements in the Global Economic Political Uncertainty Index (GEPU) and the Fear Index (VIX) on crypto currencies. Thus, it is aimed to contribute new empirical findings to the literature examining the volatility interaction between crypto currencies, global economic political uncertainty and the fear index, and to reveal important findings for investors and policy makers. For this purpose, firstly, index and crypto currencies (Bitcoin, Ethereum) return series were separated into different time scales by wavelet decomposition analysis and then examined with the DCC-GARCH method between these series. In the study, the period between GEPU, VIX and Bitcoin was April 2012-April 2024; monthly data for Ethereum between April 2016 and April 2024 used. While creating the data set of the study, these dates were determined due to data constraints for the variables.

As a result of the analysis; Findings were obtained in terms of the volatility interaction between cryptocurrencies and GEPU and VIX and four different time scales representing the short, medium and long term. The findings obtained based on raw data and disaggregated return series were evaluated separately. As a result of the analyzes carried out based on raw data; Analyzes based on raw data have obtained evidence that there is no volatility interaction between cryptocurrencies (Bitcoin, Ethereum) and GEPU and VIX returns. According to this finding, it is an indication that the shocks and dynamic conditional correlations between Bitcoin and Ethereum returns and GEPU and VIX returns in the past period have no effect on the dynamic conditional correlations between Bitcoin and Ethereum returns and GEPU and VIX returns in the current period. This result can also be expressed as there is no dynamic relationship or volatility interaction between Bitcoin and Ethereum returns and GEPU and VIX returns. When the predictions of the wavelet-based DCC-GARCH(1,1) models were evaluated, it was concluded that $\alpha + \beta < 1$ in the D_2 and D_4 time scales of Bitcoin and VIX returns was statistically significant. This finding shows that dynamic correlations fluctuate around a fixed level in the 4-8 month and 16-32 month investment cycles and have a process that tends to return to the mean. Both α coefficient estimates of DCC-GARCH(1,1) models at all time scales are statistically significant; However, the statistical significance of the β coefficients was found to be significant at the D_2 and D_4 time scales. According to these findings, it is an indicator of the persistence of volatility shocks in both 4-8 month and 16-32 month investment cycles. In addition, in the 4-8 and 16-32 month investment cycles, although $\alpha + \beta < 1$, it is very close to l (one); With VIX, it can be said that conditional volatility for Bitcoin is more likely to be permanent. In other words, there is a positive and strong relationship between returns that varies over time.

When the time scales between Bitcoin and GEPU are examined, it is concluded that there are time-varying correlations between GEPU and Bitcoin returns only in D_2 , that is, the 4-8 month investment cycle period, and that past volatility shocks and conditional correlations are effective on these correlations. In other words, volatility shocks are permanent in the 4-8 month investment cycle. In addition, in the investment cycle period of 4-8 months, although $\alpha + \beta < 1$, it is very close to l (one); With GEPU, it can be stated that conditional volatility for Bitcoin is more likely to be permanent. It has been determined that $\alpha + \beta < 1$ on the and time scales of Ethereum and GEPU returns is statistically significant. This finding shows that dynamic correlations fluctuate around a fixed level in the 4-8 month and 16-32 month investment cycles and have a process that tends to return to the mean.

Both α coefficient estimates of DCC-GARCH(1,1) models at all time scales are statistically significant; However, the statistical significance of the β coefficients was found to be significant at the D_2 and D_4 time scales. According to these findings, it is an indicator of the persistence of volatility shocks in both 4-8 month and 16-32 month investment cycles. In addition, in the 4-8 and 16-32 month investment cycles, although $\alpha + \beta < 1$, it is very close to l (one); It can be said that conditional volatility for Ethereum and GEPU is more likely to persist. On the other hand, it has been determined that there is no relationship between the fear index and the volatility shocks and dynamic correlations in the past period and the volatility shocks and dynamic conditional correlations in the current period in Ethereum returns according to different time scales.

In conclusion, these findings have practical implications for both investors and policy makers. As both the economic and political uncertainty index and the fear index are the volatility interaction between global currencies, this study shows that this affects the cryptocurrencies Bitcoin and Ethereum. Therefore, investors need to have comprehensive information about changes in the global economy and politics. Information-related policy changes should be factored into portfolio selection to avoid random market fluctuations. In addition, investors can obtain comprehensive information about the global economy and policy changes in the market, which is more turbulent and subject to sudden changes in the short term due to its unregulated structure. It is thought that these results will provide insight for both investors and policy makers.

REFERENCES

- AKDAĞ, S. (2019). VIX Korku Endeksinin Finansal Göstergeler Üzerindeki Etkisi: Türkiye Örneği. *Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 12(1): 235-256.
- AYDOĞDU, A. (2024). *Farklı Yatırım Ufuklarına Göre Kripto Para Birimlerinin Volatilite Modellemesi*, Yayınlanmamış Doktora Tezi, Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü, Denizli.
- BAKER, S. R., BLOOM, N. ve DAVİS, S. J. (2016). Measuring Economic Policy Uncertainty.
- BAKIR, E. (2021). *Covıd-19 Pandemisi Sürecinde Kripto Para Birimleri ile Ekonomik Göstergeler Arasındaki İlişki*, Yayınlanmış Yüksek Lisans Tezi, Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü, Balıkesir.
- BENHAND, F. (2013). Bul Lor Bear Markets: A Wavelet Dynamic Correlation Perspective. *Economic Modelling*, 32, 576-591.
- CoinMarketCap (2024). https://coinmarketcap.com/ (10.04.2024).
- CORBET, S., MEEGAN, A., LARKİN, C., LUCEY, B. and YAROVAYA, L. (2018). Exploring The Dynamic Relationships Between Cryptocurrencies And Other Financial Asset. *Economics Letters*, 165, 28-34.
- CORNİSH, C.R., BRETHERTON, C. S., & PERCİVAL, D. B. (2006). Maximal Overlap Wavelet Statistical Analysis With Application To Atmospheric Turbulence. *Boundary-Layer Meteorology*, 119(2),339-374.7
- CROWLEY, P. M. (2007). A Guide To Wavelets For Economists. *Journal of Economic Surveys*, 21(2), 207-267.
- DAUBECHİES, I. (1992). *Ten Lectures on Wavelets, Montpelier*, VT: Capital City Press.
- DAVİS, S. J. (2016), Global Economic Policy Uncertainty Index Data. (Erişim: 25.04.2024), http://www.policyuncertainty.com/global_monthly.html
- ENGLE, R. F. (2002). A Simple Class Of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business and Economics Statistics*, 20(3), 339-350.
- GENÇAY, R., SELÇUK, F. VE WHİTCHER, B. (2002). *An Introduction to Wavelets and Other Filtering Methods in Finance And Economics*, Academic Pres, U.S.A.
- GÖKALP, B. T. (2022). Kripto Para Piyasasının Borsa İstanbul Endeksleri Üzerindeki Etkileri. *Ekonomi Politika ve Finans Araştırmaları Dergisi*, 7(2), 481-499.
- GÜRSOY S. VE KILIÇ E. (2021), Küresel Ekonomik Politik Belirsizliğin Türkiye CDS Primi ve BİST Bankacılık Endeksi Üzerindeki Volatilite Etkileşimi: DCC-GARCH Modeli Uygulaması, *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 35(4): Sayfa: 1323-1334,
- İLTAŞ, Y. (2020), Farklı Risk Türleri ve Borsa Endeksi Arasındaki İlişki: Türkiye İçin Nedensellik Testleri, *Business & Economics Research Journal*, 11(2), 371-384.
- JAMMAZİ, R., & ALOUİ, C. (20109. Wavelet Decomposition And Regime Shifts: Assesing The Effects of Crude Oil Shock Market Returns. *Energy Policy*, 38(3), 1415-1435.
- KEVSER, M. (2023). Bitcoin as an Alternative Financial Asset Class: Relations Between Geopolitical Risk, Global Economic Political Uncertainty, And Energy Consumption. *Pamukkale Journal of Eurasian Socioeconomic Studies*, 9(2), 117-
- KHAN, K., SUN, J., DERİNDERE KOSEOGLU, S., & REHMAN, A. U. (2021). Revisiting Bitcoin Price Behavior Under Global Economic Uncertainty. *Sage Open*, *11*(3).

- LEBO, M. J., & BOX-STEFFENSMEİER, J. M. (2008). Dynamic Conditional Correlations in Political Science. *American Journal of Political Science*, 52(3), 688-704.

- LEHKONEN, H., & HEİMONEN, K. (2014). Timescale-Dependent Stock Marke Comovement: BRICS vs. Developed Markets, *Journal of Empirical Finance*, 28, 90-103.

- MASİH, M., ALZAHRANİ, M., & AL-TİTİ, O. (2010). Systematic Risk and Time Scales: New Evidence From An Application Of Wavelet Approach To The Emerging Gulf Stock Markets. *International Review of Financial Analysis*, 19(1), 10-18.

- PERCİVAL, D. W., & WALDEN, A. (2000). Wavelet Methods for Time Series Analysis. Cambridge University. San Diego, California, USA: Academic Press.

- RAMSEY, J.B., & LAMPART, C. (1998a). Decomposition of Economic Relationships by Timescale Using Wavelets. *Macroeconomic Dynamics*, 2(1), 49-71.

- RAMSEY, J.B., & LAMPART, C. (1998b). Decomposition of Economic Relationships by Timescale Using Wavelets: Expenditure and Income. *Studies in Nonlinear Dynamics and Econometrics*, 3(1), 23-42.

- RUA, A., & NUNES, L. (2009). International Comovement of Stock Market Returns Wavelet Analysis. *Journal of Empirical Finance*, 16(4), 632-639.

- SCHLEİCHER C. (2002). An Introduction to Wavelets for Economists. *Staff Working Papers*, No. 02-3, Bank of Canada.

- ŞAHİN DAĞLI, S., & ÇELİK, İ. (2022). Return and Volatility Spillover Between Energy Commodities: Evidence From the VAR-EGARCH Model. *Gaziantep University Journal of Social Sciences*, 21(4), 2187-2207.

- TARI, R. (2014). *Ekonometri*, Umuttepe Yayınları, Kocaeli.

- TUNCEL, M. B., & GÜRSOY, S. (2020). Korku Endeksi (VIX), Bitcoin Fiyatları ve Bist100 Endeksi Arasındaki Nedensellik İlişkisi Üzerine Ampirik Bir Uygulama. *Elektronik Sosyal Bilimler Dergisi*, 1999-2011.

- UYAR, U., & KANGALLI UYAR, S. (2021). Sermaye ve Altın Piyasaları Arasındaki Yayılım Etkisi: Wavelet'e Dayalı Dinamik Koşullu Korelasyon Yaklaşımı, M. Ural ve Ü. Aydın (Ed), *Finansal Ekonometri Uygulamaları Kavram- Teori-Uygulama* içinde (309-335), Seçkin Yayıncılık, Ankara.

"This page is left blank for typesetting"

HOLISTENCE publications

JAME
E-ISSN:2791-7401
Volume : 4, Issue 1, 2024
URL: https://journals.gen.tr/jame

Received: 2023-12-27
Accepted: 2024-04-25

RESEARCH ARTICLE

A Solution to Errors-in-variables Bias in Multivariate Linear Regression using Compact Genetic Algorithms

Mehmet Hakan Satman[1] **Erkin Diyarbakırlıoğlu**[2]

[1] Prof. Dr., Istanbul University, Faculty of Economics, Department of Econometrics, Türkiye, email: mhsatman@gmail.com
[2] IAE Paris-Est, University of Paris-Est - Cr´eteil, (IRG, EA 2354), Place de la Porte des Champs, France,
e-mail: erkin.d@u-pec.fr

Abstract

We address the classical errors-in-variables (EIV) problem in multivariate linear regression with N dependent variables where each left-hand-side variable is a function of a common predictor X subject to measurement error. Our contribution consists in employing the remaining $N-1$ regressions as *extra information* to obtain a filtered version of the mismeasured series X. We test the performance of our approach using simulations whereby we control for different cases like low vs. high R^2 models, small vs. large sample or small vs. large measurement error variances. The results suggest that the multivariate-Compact Genetic Algorithm (mCGA) approach yields estimates with lower mean-square-errors (MSEs). The MSEs are decreasing as the number of dependent variables increases. When there is no measurement error, our method gives results similar to those that would have been obtained by ordinary least-squares.

Keywords: Multivariate regression, Errors-in-variables, Compact genetic algorithms

JEL codes: B23, C13, C63, C61

Citation: SATMAN, M.H. & DİYARBAKIRLIOĞLU, E. (2024). A Solution to Errors-in-variables Bias in Multivariate Linear Regression using Compact Genetic Algorithms. *Journal of Applied Microeconometrics (JAME)*. 4(1),31-64, DOI: 10.53753/jame.2293

Corresponding Author:
Mehmet Hakan Satman
E-mail: mhsatman@gmail.com

1 Introduction

Errors-in-variables (EIV) occur when the observations of one or more variables in a regression model do not match their true values and, consequently, contain a measurement error. Basically, EIV in the left and/or right-hand-side variables in a statistical model can be read as *the observation equals the true values plus measurement error* with side effects ranging from "mild" to "severe" to the researcher. The econometric treatment of EIV is at the origin of a rich yet inconclusive literature going back as far as to Frisch, Berkson or Durbin's pioneering works (Frisch, 1934; Berkson, 1950; Durbin, 1954). More recent and exhaustive treatments of the topic include, among others, Feng et al. (2020), Racicot (2015), Chen et al. (2011), Buonaccorsi (2010), Davidson and MacKinnon (2004, ch. 8), Hausman (2001), Bound et al. (2001), Hyslop and Imbens (2001), Cheng and Van Ness (1999), Dagenais and Dagenais (1997), Griliches (1987) or Fuller (1987), only to cite a few. This paper's primary objective is to contribute to this literature by addressing the classical EIV model in the context of a linear regression setting with multiple dependent variables where a single independent variable X, possibly measured with error, is linked to N dependent variables $Y_1, ..., Y_N$.

Specifically, we extend the work of Satman and Diyarbakirlioglu (2015) who develop a modern approach to deal with EIV that requires no extra information nor additional data to mitigate the bias generated by the measurement error in the independent variable. We consider this feature of our approach as the central block that marks off our work from previous studies in the field. A detailed exposition of existing methods would inevitably extend the paper's scope beyond acceptable limits, so we outline a concise discussion in due course.

A first, and rather naive, way of addressing the measurement error consists in simply ignoring the problem by admitting that it is a difficult one to solve and extra data may not be available to the researcher. A second approach, known as *Berkson's approach* (Berkson, 1950) considers the observed X values as predetermined given, say, a controlled scientific study or under a laboratory setting. Then, one would consider the observed values of X swinging around their true equivalents due to measurement error. Under this setup, it is then reasonable to assume that the measurement error is no longer correlated with the observed values of the independent variable, which enables one to show that the traditional least-squares estimator of the slope remains still unbiased even when X is mismeasured (Durbin, 1954, p. 24-25). When it comes to social and economic phenomena, however, such controlled experiment settings do not truly exist outside the laboratory.

Another approach relies on correcting the bias in the estimates assuming that the variance of the measurement error in the predictor or that of the unobserved predictor is known. This would then make it possible to derive unbiased estimates of model parameters using the *signal-to-total* variance, typically known as the *reliability ratio*. (Fuller, 1987, p. 5-6) gives a list of some situations where the reliability ratio can be considered as known like IQ test scores. Such situations where one would plausibly assume that the reliability ratio is known are however mostly limited to survey studies in which the data about a particular feature of a set of respondents are obtained over repeated studies of the same nature over time and space.

Given the shortcomings associated with the approaches described briefly above and to the extent that EIV naturally induces a specification error in regression models, instrumental-variables estimation of EIV models constitutes the central prescription to address the issue. The main idea of IV-based processing of EIV consists in using instruments correlated with the true but unobserved values of the predictor *and* uncorrelated with the measurement error, see, among others, Fuller (1987), Davidson and MacKinnon (2004), Carroll et al. (2006) or Wooldridge (2010) for further developments on the IV-estimation of EIV models. As noted by (Buonaccorsi, 2010, p. 130), the instruments are supposed to carry independent information about the mismeasured predictor which can be used to obtain estimates of the coefficients. That being said, the validity condition of good instruments just stated previously is ironically the unique but also the most critical potential drawback associated with the IV-estimation of EIV models to the extent that poor instruments lead to even more serious consequences (Wooldridge, 2013, p. 499).

The bottomline is that existing approaches commonly require additional information in the form of either better data or valid instruments (Shalabh et al., 2010, p. 718). While there are nice examples where the researcher comes up with an ingenious solution to overcome the impact of a badly measured variable like the studies on the estimates of the economic return to schooling (Ashenfelter and Krueger, 1994; Harmon and Walker, 1995), such additional information may not be available in other situations or there may be no consensus in the field as to what makes an instrument a good one (Klepper and Leamer (1984, p. 163), Dagenais and Dagenais (1997, p. 194)). The approach proposed here is free from such considerations. It does not require any out-of-the-system information about the data-generating process to mitigate the EIV problem. This is a key feature in that the unique *extra* information required lies in the additional dependent variables of the system. We conceive the mismeasured variable $X^\star$ as consisting of two blocks, one deterministic and the other stochastic where the deterministic part refers to the true but unobserved portion of it. We then devise an optimization problem that minimizes the squared deviations from the expectation of the response variable conditional on the estimated values of the mismeasured predictor. The latter variable in turn is the result of an auxiliary dummy regression of the initial variable subject to measurement error. The key difference with the initial Satman and Diyarbakirlioglu (2015) study is that we consider the case where several dependent variables are connected to the same independent variable, potentially measured with error.[1]

We employ numerical experiments to gain insights into the performance of our method. Following a parsimonious strategy to devise the simulations, we consider 36 different configurations to control for features like the relative ratio of regression vs. measurement error variances, low vs. high R^2 models, the sample size and the number of dependent variables. For each configuration, we repeat the estimations 1,000 times. We report the bias, the variance and the mean-square-error of the coefficient estimates for the first dependent variable.[2] The results are promising. The algorithm does capture and correct the bias due to the measurement error in the independent variable, which, when ignored, distorts seriously the parameter estimates. The bias in the slope tends to vanish as the sample size or the number of left-hand-side variables increases. This comes with some cost in the increase of the estimator's variance but overall the increase in the variance is largely offset by the decrease in the bias. This, collectively, yields much smaller MSEs, giving further credit to our approach.

The paper is organized as follows. After a brief discussion of the classical EIV model, section 2 describes our algorithm and gives a short discussion some of its important features. Section 3 presents the simulation results. In section 4,

[1] One should also highlight that the solution presented in Satman and Diyarbakirlioglu (2015) can be considered as a specific case of the setup we develop therein with the number of dependent variables set to 1.

[2] This choice is motivated by the fact that it would be nearly impossible to report every single intercept and slope estimate for each Y_i in the model as this would make the size of the paper cross the acceptable limits. That being said, we saved the entire output from each set of estimations. We have also performed the procedure for 18 additional configurations where we controlled for the performance of our approach when there is no measurement error in the independent variable. For sake of brevity, we do not report the results of these additional simulations in the paper. These results are available upon request.

we give a simple example to illustrate the implementation of our algorithm. Section 5 concludes.

2 Methodology

2.1 The multivariate CGA Algorithm

We first give a sketch of the consequences of the classical EIV model and present our methodology afterwards.

Consider the population model $Y_t^\star = \beta_0 + \beta_1 X_t^\star + \epsilon_t$ for $t = 1, ..., T$ with $\epsilon \sim iid(0, \sigma_e^2)$. The classical errors-in-variables (EIV) model is introduced by assuming that the observations on $Y^\star$ and/or $X^\star$ are recorded with error as $Y_t = Y_t^\star + \nu_t$ and/or $X_t = X_t^\star + \delta_t$ where ν and δ are observation, or measurement errors on Y and X respectively.[3]. While the cost of ν is limited to an inflated variance of the regression error, matters are different when it comes to δ.[4] Assuming $Var(\nu) = 0$, $Var(\delta) > 0$, $E(X^\star\delta) = 0$ and $E(\delta\epsilon) = 0$[5], simple algebra shows that the model can now be expressed as $Y_t = \beta_0 + \beta_1 X_t + \omega_t$ where $\omega_t = \epsilon_t - \beta_1\delta_t$.

Thus, we obtain a *composite* regression error and the predictor X becomes correlated with the new disturbance term as $Cov(X_t, \omega_t) = Cov((X_t^\star + \delta_t), (\epsilon_t - \beta_1\delta_t)) = -\beta_1 Var(\delta)$. This implies that the least-squares estimate will be biased and inconsistent even in large samples.[6] In addition, given the probability limit plim $\widehat{\beta}_1^{LS} = \beta_1 + \frac{Cov(X_t,\omega_t)}{Var(X_t)}$, which is also commonly expressed as plim $\widehat{\beta}_1^{LS} = \beta_1 \left(\frac{Var(X_t^\star)}{Var(X_t^\star)+Var(\delta_t)}\right)$, it can be seen that the slope

[3] Typically, these equations read “the observation is the sum of the true value plus *measurement error*”

[4] If $Var(\nu) > 0$ and $Var(\delta) = 0$, the model can be rewritten as $Y_t = \beta_0 + \beta_1 X_t^\star + (\epsilon_t + \nu_t)$. The new disturbance term is $\epsilon + \nu$. The least-squares estimates of parameters will still be unbiased.

[5] One last assumption holds that the measurement error, by definition, has zero mean, $E(\delta) = 0$.

[6] To see the non-zero covariance between X and ω, note that $E(X_t) = E(X_t^\star + \delta_t) = X_t^\star$ because $E(\delta_t) = 0$. Next, substituting $\omega_t = \epsilon_t - \beta_1\delta_t$ back into $Cov(X_t, \omega_t) = E\left[(X_t^\star + \delta_t - X^\star)(\epsilon_t - \beta_1\delta_t)\right]$ and developing the terms, we obtain $Cov(X_t, \omega_t) = -\beta_1 Var(\delta)$.

estimate is downwards biased as long as $Var(X_t^\star)+Var(\delta_t) > Var(X_t^\star)$.[7] The bias in $\widehat{\beta}$ is known as the *least-squares* attenuation, which Hausman (2001) refers to as the *iron law of econometrics* – "the magnitude of the estimate is usually smaller than expected". The attenuation in $\widehat{\beta}$ also suggests that the bias gets worse when $Var(\delta_t)$ increases relative to $Var(X_t^\star)$.[8] Finally, when there is more than one predictor subject to error, it is no longer possible to derive exact formulas to express neither the sign nor the magnitude of the bias in the slope coefficients because the measurement error on a particular X_t spills over to other model parameters, raising a further puzzling issue, which Cragg (1994) qualifies as the *contamination effect.*

We now turn to our approach and extend the previous setup to accommodate for N dependent variables specified as a function of the same predictor variable. We assume the population model is $\mathbf{Y} = \mathbf{1}_T\boldsymbol{\beta}_0^\top + \mathbf{X}^\star\boldsymbol{\beta}_1^\top + \boldsymbol{\epsilon}$.[9] The true values of $\mathbf{X}$ are not directly given but observed as $\mathbf{X} = \mathbf{X}^\star + \boldsymbol{\delta}$ where $\boldsymbol{\delta}$ is a $T \times 1$ vector of measurement errors. The multivariate EIV model can be rewritten as $\mathbf{Y} = \mathbf{1}_T\boldsymbol{\beta}_0^\top + \mathbf{X}\boldsymbol{\beta}_1^\top + \boldsymbol{\omega}$ where $\boldsymbol{\omega} = \boldsymbol{\epsilon} - \boldsymbol{\delta}\boldsymbol{\beta}_1^\top$ is the $T \times N$ matrix of composite error terms, which make the least-squares estimation of the N slope estimates inconsistent and biased in the same way it does when $N = 1$. To describe how our algorithm works, consider the first two equations of the system that relates the first and second dependent variables Y_{tj}, $j = 1, 2$ to X_t:

$$Y_{t1} = \beta_{01} + \beta_{11}X_t + \omega_{t1}$$

$$Y_{t2} = \beta_{02} + \beta_{12}X_t + \omega_{t2}$$

[7] Consistent estimation of the slope using generalized least-squares is actually possible if the value of the *reliability ratio* $\lambda = Var(X^*)/\left(Var(X^*) + Var(\delta)\right)$ is known. This is however a big "if" because the true value of the reliability ratio is also unknown outside controlled experiment settings (Buonaccorsi, 2010).

[8] The results of our simulations also highlight this fact whereby we pinpoint the case of a high ratio of measurement error variance to independent variable variance.

[9] $\mathbf{Y}$ is a $T \times N$ matrix that contains T observations for N dependent variables, $\mathbf{X}^\star$ is a T-vector of the observations on the true values of the independent variable, $\mathbf{1}$ is a conforming vector of ones, $\boldsymbol{\beta}_0$ is a N-vector of intercepts, $\boldsymbol{\beta}_1$ is a N-vector of slopes, and $\boldsymbol{\epsilon}$ is a $T \times N$ matrix of residuals.

The objective is to estimate the parameters β_{01} and β_{11} for the first equation (as well as σ^2_ω), but also the parameters β_{02} and β_{12} for the second equation too, and so on for any additional Y. The departure point of the extension we propose in this paper relative to the original approach developed in Satman and Diyarbakirlioglu (2015) consists in employing the additional variables Y_{ti} to obtain a new series $\widehat{X}_t^{mCGA}$ for the regressor that can be seen as a *filtered* version of the true yet unobserved values $X_t^\star$. This can be achieved by running the following auxiliary regression of the observed X_t on a set of m dummy variables as,

$$X_t = \alpha_0 + \alpha_1 D_{t1} + \cdots + \alpha_m D_{tm} + \eta_t \tag{1}$$

where α_j are unknown parameters that must be estimated, D_{tj} are $j = 1, ..., m$ dummy variables and η_t are regression residuals. Just like any other regression model one would conceive, this auxiliary regression breaks down the observed series X_t into two components, one deterministic and one random. By construction, the stochastic part η is an estimate of the measurement error δ while the deterministic part represents the series $\widehat{X}_t^{mCGA}$. With no closed-form solution available, the fitted coefficients $\widehat{\alpha}_j$ are devised as solution to the following problem,

$$\underset{\{D_1,\ldots,D_m\}}{\text{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{ti} - \left(\widehat{\beta}_{0i} + \widehat{\beta}_{1i} \widehat{X}_t^{mCGA} \right) \right)^2 \tag{2}$$

where $\widehat{X}_t^{mCGA}$ are themselves the fitted values of the original variable obtained from the auxiliary regression as,

$$\widehat{X}_t^{mCGA} = \widehat{\alpha}_0 + \widehat{\alpha}_1 D_{t1} + \cdots + \widehat{\alpha}_m D_{tm} \tag{3}$$

Finally, the series $\widehat{X}_t^{mCGA}$ is plugged back into the system to estimate Y_{ti} as,

$$\begin{aligned}
\widehat{Y}_{t1} &= \widehat{\beta}_{01} + \widehat{\beta}_{11}\widehat{X}_t^{mCGA} \\
\widehat{Y}_{t2} &= \widehat{\beta}_{02} + \widehat{\beta}_{12}\widehat{X}_t^{mCGA} \\
\vdots &\quad \vdots \\
\widehat{Y}_{tN} &= \widehat{\beta}_{0N} + \widehat{\beta}_{1N}\widehat{X}_t^{mCGA} \qquad (4)
\end{aligned}$$

for each $i = 1, ..., N$. Equation (2) defines a quadratic objective function subject to the constraint defined in equation (3). With T observations and $T \times m$ unknown binary values[10], the problem admits theoretically an infinite number of solutions for there is no explicit rules about the appropriate number of dummies that must be used. In their original paper, Satman and Diyarbakirlioglu (2015) address this issue by observing the behaviour of the estimated intercept and slope coefficients. They note that the MSE's of the estimates tend to stabilize about $m = 10$. We follow the same empirical rule in this paper too and use 10 as the default value of this parameter. Besides, even if the number of dummies was known, the results of the algorithm should still be seen as *approximations* sharing, nonetheless, the important feature of systematically smaller MSE's for the estimated regression parameters.[11]

2.2 Discussion

Having set up the mechanics of our approach, we briefly discuss some of its main building blocks.

[10] Recall that m is the number of dummy variables in the auxiliary regression.

[11] One should bear this feature of our method in mind: The procedure does not yield a single *exact* output, the results are likely to vary, at least marginally, from one iteration to another. That does not however mean that the algorithm does not converge, so we conceive these *approximations* as *solutions* of the system.

First, unlike the mainstream literature on EIV models, we conjecture that the additional information to mitigate the EIV bias can be found within the relationship between the set of N dependent variables and the predictor X. One should also underline that the procedure does not require further assumptions about the stochastic behaviour of X, such as its distributional properties. That is one of the key features of the approach initially adopted by Satman and Diyarbakirlioglu (2015), which we aim to develop further in this work, to the extent that standard methods generally require outside information to address the EIV problem. However, as pointed out by (Buonaccorsi, 2010, p. 4-5), getting such extra information either in the form of better data or instrumental variables that satisfy many conditions can be difficult.

Second, we explain briefly the reason why we implement a (compact) genetic algorithm-based solution. Equations (2) and (3) form together a two-stage discrete optimization problem whose objective is to minimize the squared deviations from the conditional expectation of the independent variable on a set of dummy variables and model parameters. Regression of the error-prone variable onto these dummies aims to break down this variable into a clean, but unobserved component and another one that captures the measurement error in X. Since the decision variables of the optimization problem take exclusively binary values, e.g. $D_{tm} \in \{0, 1\}$, a genetic algorithm (GA) happens to be one natural solver to estimate the dummy coefficients α of the auxiliary regression. Developed by pioneering studies like Holland (1975), Holland (1987) or Goldberg (1989), among others, a GA mimics the process of natural selection with, consequently, a related vocabulary borrowing extensively from the Theory of Evolution. A typical GA starts by encoding an array of randomly selected candidate solutions in binary forms, assimilated to *chromosomes*, each member of a larger *population*. The chances a chromosome survives for mating with another one to generate an *offspring* is determined by a *fitness value*, which is a score associated with an objective function. Iterations continue until no incremental improvement is obtained in terms of the fitness value.

A potential issue associated with GAs concerns the computational difficulties associated with the optimization of the objective function. This is where Compact Genetic Algorithms (CGAs) may be of practical help as they are designed to overcome the issue of computational memory one would face when working with a GA (Harik et al., 2006). Although one would not assert that CGAs are superior to GAs in reaching the global optimum, they represent several advantages. Specifically, in a CGA, candidate solutions are sampled from a given population using a probability vector rather than screening the entire population. The number of iterations is defined with respect to the the population size (Harik et al., 1999). The absence of genetic operators and the sampling strategy employed by a CGA make it a member of Estimation of Distribution Algorithms (EDA) as it always converges to a probability vector through iterations (Pelikan et al., 2002; Baluja, 1994; Larranaga, 2002). Therefore, our choice in implementing the CGA is simply motivated by the fact that the algorithm provides a suitable method to solve the discrete optimization problem defined in equation (3), yet it should be acknowledged that another optimizer handling a similar problem would also be used instead.

Finally, we present some practical, but equally important, aspects of our methodology.[12] Given a set of T observations on $i = 1, ..., N$ dependent variables Y_i and one independent variable X observed with some error δ, the procedure is initialized by setting two user-defined parameters; namely, (1) the number of dummy variables m used in the auxiliary regression specified in equation (3), and (2) the population size. Although there are no specific

[12] We provide in the appendix a pseudo-code of our entire algorithm and an R package (Satman and Diyarbakirlioglu, 2022) including all necessary functions to perform the calculations is readily available on CRAN repositories.

guidelines concerning an adequate value for m, Satman and Diyarbakirlioglu (2015) show using simulations that the mean-square-errors of the slope $\widehat{\beta}_1^{CGA}$ and intercept $\widehat{\beta}_0^{CGA}$ estimators stabilize around $m = 10$.[13] For a given m, the iterations begin with a *probability vector* that represents, to speak CGA, a *chromosome*. For example, a 4-m length chromosome like,

$$P = [0.8, 0.1, 0.7, 0.2]$$

tells that the probability of getting the first dummy equal to 1 is 0.8, the probability of $D_2 = 1$ is 0.1, and so on.[14] Accordingly, given the P in this example, sampling a chromosome like $C = [1, 0, 1, 0]$ is much more likely than sampling another chromosome like $C' = [0, 1, 0, 1]$. Once the number of dummies is chosen, which we set to 10, iterations begin with a probability vector whose elements are initially all equal to 0.5, guaranteeing that no specific dummy coefficient is favoured relative to others. In the next step, the procedure samples two *parents* using the initial P, say C_1 and C_2. The *winner* is the one with the lowest score of the *cost function*, which is specified as the sum of the squared residuals of the corresponding dummy regression. Once the winner C is determined, the vector P is updated using the formula,

$$P_{i+1} = \begin{cases} P_i + \frac{1}{\text{pop. size}} & \text{if } C_i^{winner} = 1 \\ P_i - \frac{1}{\text{pop. size}} & \text{if } C_i^{winner} = 0 \end{cases} \tag{5}$$

Given the new P_{i+1}, the process moves forward by sampling new parents, generating new offsprings and updating thereby P_i. Iterations continue until all

[13] See Satman and Diyarbakirlioglu (2015), figure 1, p. 3225. We also follow the same empirical rule suggested by the authors in the original paper and set $m = 10$ in our applications.

[14] The term *probability vector* should then not be understood in the sense the elements of the vector must sum up to 1. Instead, each element of P defines the probability that the corresponding dummy variable to be equal to 1.

elements of the vector P take either the value of 1 or 0. Note that the *population size* is there for updating iteratively to the probability vector until the stability condition for the auxiliary dummy variables regression is obtained.[15]

3 Simulations

3.1 Setup

We investigate the statistical properties of our approach by Monte Carlo simulations. We specify our data-generating process as follows:

$$
\begin{aligned}
Y_{ti} &= 5 + 5X_t + \epsilon_{ti} \\
X_t &= X_t^\star + \delta_t \\
\epsilon_t &\sim iid\ N\left(0, \sigma_\epsilon^2\right) \\
\delta_t &\sim iid\ N\left(0, \sigma_\delta^2\right)
\end{aligned}
$$

The index $t = 1, \ldots, T$ shows the sample size and $i = 1, ..., N$ the number of left-hand-side variables. Each configuration is described by four parameters: (1) The number of left-hand-side variables N, (2) the sample size T, (3) The regression error variance σ_ϵ^2 and, (4) the measurement error variance σ_δ^2.

We choose three different values for $N \in \{2, 5, 25\}$ to construct the multivariate regression setting and three different sample sizes as $T \in \{30, 50, 100\}$. The measurement error δ is introduced as $X = X^\star + \delta$. ϵ and δ are both generated as *iid* normal random variables with zero-mean and constant variance as $\epsilon \sim N(0, \sigma_\epsilon^2)$ and $\delta \sim N(0, \sigma_\delta^2)$. Regarding the "regression error ϵ & measurement error δ" pairs, we distinguish four configurations as we set $\sigma_\epsilon \in \{1, 3\}$

[15] The choice of the population size takes into consideration the trade-off between the *convergence speed* vs. the risk of a local optimum trap. Again, we follow the recommendations of Satman and Diyarbakirlioglu (2015) who suggest that the population size should be 20 or higher. That is said, the authors also note that beyond this limit, the population size has negligible effect on the results. For the record, this parameter is set to 40 in our applications.

together with $\sigma_\delta \in \{0.5, 0.9\}$.[16] Therefore, we distinguish between small vs. large samples as well as low vs. high R^2 models by considering different pairs of regression error vs. measurement error variances, σ_ϵ vs. σ_δ. This allows to study the performance of our method with small vs. large attenuation bias.

Collectively, there are 36 different configurations, which we run 1,000 times each. We thus report the results using a total of 36,000 simulated regressions. We also repeat the experiment for 18 additional configurations[17] whereby we control for the case with no measurement error in X and keep other parameters constant. Our objective is to verify the accuracy of the process in the idealized case and to compare the least-squares with the mCGA method. We find no significant difference between the statistical properties of the two methods. The mCGA therefore conveys no erroneous signal in the absence of measurement error.

3.2 Simulation results

We report our results in Tables 1 to 3. These tables show the results by the number of dependent variables, i.e. $N = 2$, 5, and 25, respectively. We calculate for each configuration the parameter bias as $E(\widehat{\theta}) - \theta$, the variance $Var(\widehat{\theta})$ and the $MSE = \left(E(\widehat{\theta}) - \theta\right)^2 + Var(\widehat{\theta})$ of the intercept β_0 and slope β_1 estimates. We also provide two additional tables in which we organize the results by $\sigma_\epsilon \& \sigma_\delta$ pairs and for increasing number of observations T to enable a complementary reading of our numerical experiments. These are given in Tables 4 and 5. As a supplement, we also provide a graphical summary of part of the output in Figures 1 and 2 where we show the bias and the MSE scores of the slope estimates, broken down by the number of dependent variables.

We can make several observations on the basis of our numerical experiment.

[16] We also consider the case with no measurement error by setting $\sigma_\delta = 0$. For sake of brevity, we do not report the results of these configurations, which are available upon request.
[17] The results are available upon request.

First, we observe, regardless of the configuration, that the least-squares estimate of the slope suffers from the attenuation bias when the predictor is subject to measurement error. For example, when we set $\sigma_\epsilon = 1$ and $\sigma_\delta = 0.5$, one would expect that the least-squares estimate of the slope to be biased downwards by 20% relative to the true value of the parameter, which implies that β_1 set initially to 5 will be cut down to 4. This observation holds indeed for the case $(\sigma_\epsilon = 1, \sigma_\delta = 0.5)$ in Tables 1 to 3 regardless of the sample size chosen. The bias in β_1 is even more pronounced when the variability of the measurement error increases relative to that of the regression error.

Second, the performance of our method in mitigating the attenuation bias in the slope is noticeable. In some cases, especially for the $\sigma_\epsilon = 1$ & $\sigma_\delta = 0.5$ pair, the algorithm comes up with an estimate of the slope fairly close to the true value of the parameter. In addition, we observe, as one would expect from our setup, even better-behaved results for the bias in β_1 as we increase the number of dependent variables across Tables 2 and 3. That is said, the decrease in the bias of the slope estimate is not homogeneous as for larger values of the regression error variance. To sum up, the simulations yield a systematically lower bias of the CGA-estimate of the slope $\widehat{\beta}_1^{mCGA}$ relative to that the least-squares estimate $\widehat{\beta}_1^{LS}$ for all configurations.

Third, we look at the variance and mean-square errors (MSE) of the estimates. Overall, we note that the variance of the estimates remains stable across different simulation configurations; while the number of left-hand-side variables has seemingly no effect on the variance, the sample size appears to significantly flatten the variance of the mCGA estimates within a given simulation configuration. Given the decrease in the bias, this results in lower MSE associated with $\widehat{\beta}_1^{mCGA}$, as suggested in Tables 2 to 3 for any configuration one would consider. For a given number of dependent variables N and the σ_ϵ

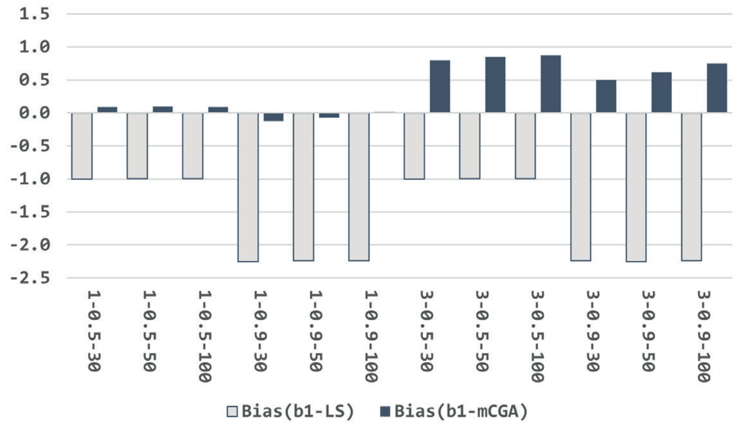
& σ_δ pair, we observe that the MSEs decrease systematically as the sample size increases. This can be easily observed in Table 4. The same observations also hold for the MSE values of the intercept estimates. On the other hand, when we consider the results for the intercept in Tables 1 to 3 and 5, we note that, as expected, the least-squares estimator tends to outperform the mCGA as $\widehat{\beta}_0^{LS}$ remains unbiased with minimum variance even with measurement error. $\widehat{\beta}_0^{mCGA}$, however, bears bias and variance values comparable to those of the least-squares. For example, in the first row of table 3 with $\sigma_\epsilon = 1$ and $\sigma_\delta = 0.5$, we read the bias in $\widehat{\beta}_0^{mCGA}$ as 0.0072 while it is -0.0032 for the LS. In addition, the MSEs of $\widehat{\beta}_0$'s are insensitive to the simulation configurations, seemingly independent of the number of dependent variables and decreasing as the sample size increases. We also note that an increase in the measurement error standard deviation and that of the disturbances tend to degrade the statistical properties of the intercept estimator while an increase in σ_δ has more destructive effects than an increase in σ_ϵ.

To conclude this section, we also give a graphical summary of the message carried out by our numerical experiments. Specifically, in figures 1 and 2, we show using bar charts how the bias and the MSE of estimated β_1's change when one applies the m-CGA estimator (dark bars) relative to least-squares (grey bars). We consider three panels to distinguish the three different values we chose for the number of dependent variables N. The x-axis labels consist of three consecutive numbers that define a given simulation configuration, namely (1) the regression error standard deviation σ_ϵ, (2) the measurement error standard deviation σ_δ, and (3) the sample size T.

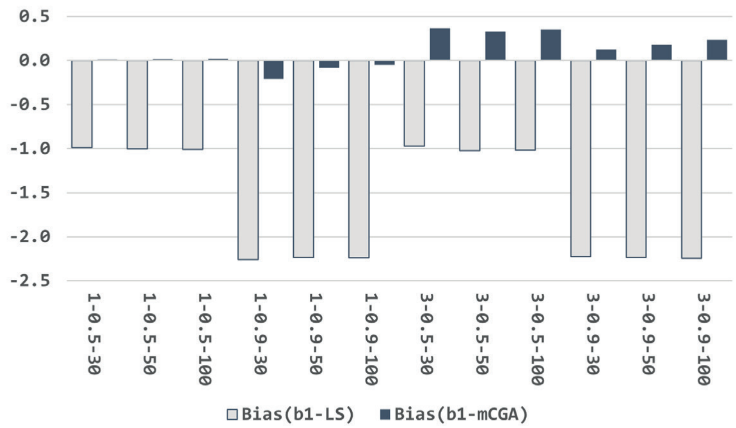
The bars are of the same length regardless of these values for the bias and MSE scores associated with the LS estimates. In a nutshell, the shorter the bars, the better the results, which is the case for every configuration we consider in terms of both bias and MSE of the estimates: The CGA estimates of

Figure 1: Comparing Bias($\widehat{\beta}_1$), OLS vs. mCGA

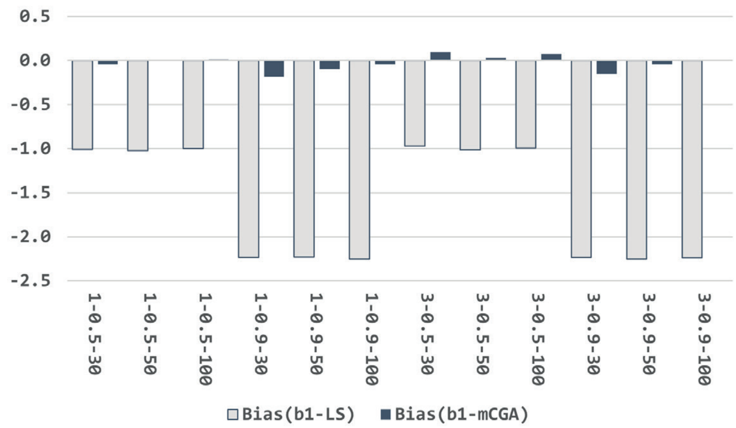
(a) Panel A: N = 2



(b) Panel B: N = 5



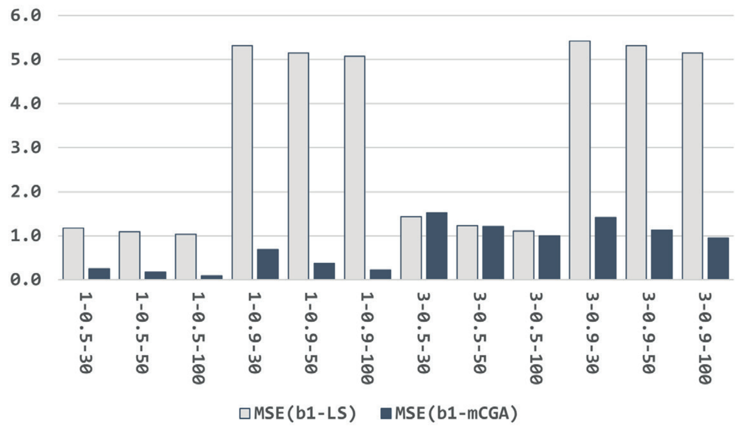
(c) Panel C: N = 25



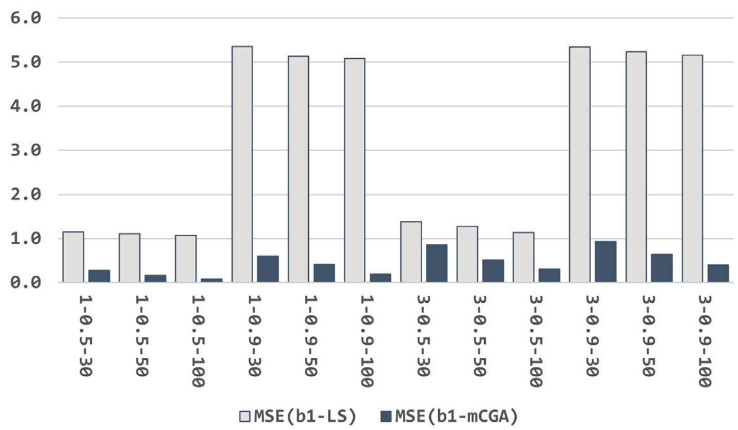
the regression slope have better statistical properties than those of their least-squares equivalents. The downward bias caused by the measurement error is

Figure 2: Comparing MSE($\widehat{\beta}_1$), OLS vs. mCGA

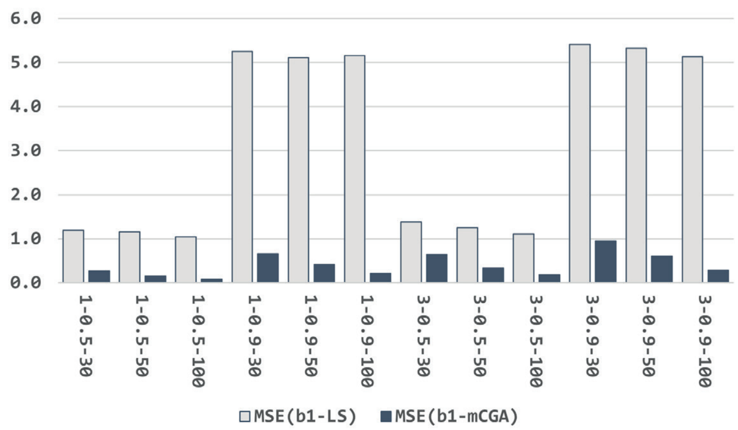
(a) Panel A: N = 2



(b) Panel B: N = 5



(c) Panel C: N = 25



noticeable based on the plots on the left-side of the figures. The CGA estimator is on the other hand successful in pulling the estimate back to its original value in such a way that Bias $\left(\widehat{\beta}_1\right)$ nearly disappears as suggested in panel C of figure 1 where we consider the case with 25 left-hand-side variables in the multivariate regression.

4 Empirical illustration

We provide a simple empirical illustration of our approach. We prefer an idealized setup for ease of tractability of the results in a multivariate multiple regression model with two response variables and three predictor variables, one of which is measured with error. For $t = 1, ..., 25$ observations and the ith response variable, $i = 1, 2$, we consider the following data generating process:

$$\begin{aligned} Y_{ti} &= \alpha + \beta X_t^\star + \gamma_1 W_{t1} + \gamma_2 W_{t2} + \epsilon_{ti} \\ \epsilon_i &\sim iid\ N(0,1) \\ X^\star, W_1, W_2 &\sim iid\ N(0,1) \end{aligned}$$

Table 6 shows our artificial dataset. The model parameters α, β, γ_1 and γ_2 are all equal to 5. Therefore, the population regression function for the first response variable is $Y_{t1} = 5 + 5X_t^\star + 5W_{t1} + 5W_{t2} + \epsilon_{t1}$. Then, we introduce the measurement error in $X^\star$ using $X_t = X_t^\star + \delta_t$ where $\delta \sim iid\ N\left(0, 0.5^2\right)$.

We focus on the β coefficient associated with the variable X and fit the following regressions to analyze the behaviour of the coefficient β in,

$$Y_{t1} = \begin{cases} \alpha_i + \beta X_t^\star + \gamma_1 W_{t1} + \gamma_2 W_{t2} + \epsilon_{t1} & \text{Model 1: No EIV} \\ \alpha_i + \beta X_t + \gamma_1 W_{t1} + \gamma_2 W_{t2} + \omega_{t1} & \text{Model 2: EIV} \\ \alpha_i + \beta X_t^{mCGA} + \gamma_1 W_{t1} + \gamma_2 W_{t2} + u_{t1} & \text{Model 3: mCGA} \end{cases}$$

The first model is the initial case with no measurement error whereby the least-squares method is expected to yield a BLUE estimator of β. The second model involves the errors-in-variables case where we expect an attenuation bias by 80% in β. The true beta being equal to 5, the fitted beta with error in X should be close to $\beta \times \left(1/(1+0.5^2)\right) = 4$.[18] Model 3 shows the case where we implement our method to mitigate the errors-in-variables bias. Table 7 summarizes the estimation results.

The message of the example is conspicuous. With no error in $X^\star$, on the first column of table 7, the OLS yields virtually "perfect" results as long as we consider the idealized where several assumptions of the estimator hold within the simulation setting from the very beginning. When we add the measurement error δ and run the model using X, as reported by model 2, the point estimate of β is downsized, unsurprisingly, by more than 20%, going down from 5.048 to 3.931, with a standard error nearly twice as much as the one found by the OLS. Finally, the mCGA estimator is remarkably successful as it pulls the β associated with the mismeasured variable back to 4.662.

Additional practical and important observations concern the pairwise relationships between the variables of interest. We mentioned earlier in section **??** that the principal feature of the algorithm we devise consists in filtering out the variable X into two components as $X = \widehat{X}^{mCGA} + \hat{\eta}$: The random component stands for the estimate of the additive measurement error δ while the deterministic component $\widehat{X}^{mCGA}$ matches the *fitted* X, which we use in the second-stage regressions.[19] These two parts must then be uncorrelated. This is indeed the case: The sample correlation between the measurement error δ and the fitted errors is $Cor(\delta, \hat{\eta}) = 0.8786$, suggesting that the algorithm

[18] The calculation is possible thanks to the knowledge about the measurement error and regression error variances.

[19] There are other instances in the EIV literature following a similar *two-stage path* like the one we introduce here. See, among others, Dagenais and Dagenais (1997), Racicot (2015).

comes up with an accurate estimate of the measurement error. In addition, by assumptions of the classical EIV model, we expect the measurement error δ to be uncorrelated with the true values of the predictor $X^{\star}$. The weak sample correlation between the two series in our data validates this insight: $Cor(\delta, X^{\star}) = 0.2007$. Finally, and above all else, the filtered series $\widehat{X}^{mCGA}$ has a very strong correlation with the true values $X^{\star}$ (assumed unobserved). The correlation between $\widehat{X}^{mCGA}$ and $X^{\star}$ is 0.9726. In words, the method comes up with a *clean series* for the variable of interest, by providing very close to the true but unobserved series.

5 Conclusion

This paper addresses the classical errors-in-variables problem in multivariate linear regression by introducing a compact genetic algorithm-based estimator designed to mitigate the EIV bias. We build on the original work by Satman and Diyarbakirlioglu (2015). The authors developed a framework that considers the measurement error problem within a constrained convex optimization setting and generates a cleaner version of the error-prone regressor with no outside information. This paper extends their idea in a multivariate regression system involving N response variables, where each variable is a function of the same regressor and, doing so, aims to take advantage of the additional information provided by the $N-1$ variables to obtain better-behaved estimates of model coefficients.

In the same spirit as the original paper, our approach consists of a two-stage optimization process in which the first stage comes up with a filtered version of the independent variable through an auxiliary dummy-variables regression. The new series is then plugged back into the initial model in the second stage to mitigate the EIV problem. We perform extensive simulation analyses to assess our approach. We consider several control parameters like the sample size, the number of dependent variables of the multivariate regression or the regression vs. measurement error variances. We also provide a simple empirical application of our method again using simulated data to further highlight the accuracy of our approach. To summarize, the results overall suggest that the inclusion of additional response variables as extra information reduces the bias at the expense of a relatively tolerable increase in the variance. That is said, the increase in the variance is largely offset as we observe systematically smaller MSE's in all simulation configurations, endorsing the performance of our approach.

There are several options for future studies. A direct extension would focus on an in-depth investigation of the statistical properties of our estimator. Although the slight increase in the variance is substantially offset by lower bias in the coefficient estimates, yielding, collectively, systematically lower mean-square-errors, studying other features of our method like the consistency, decision error probabilities or robustness is equally desirable. Therefore, we consider such a simulation-driven study as a starting point for future work to provide further credit to the framework we aim to develop. Another avenue for future work concerns the empirical ground so that one would check the CGA estimator in action with real data.[20] There are of course several instances in different disciplines in which the model involves a linear relationship between several response variables each function of the same set of regressors. For example, a particularly interesting case in financial economics is the so-called *factor pricing models* where many dependent variables, i.e. returns on a set of assets or portfolios, are modelled as a linear function of a given set of independent variables, i.e. risk factors. In a recent study conducted by

[20] As a matter of fact, the main issue related to empirical work is that it is rarely possible to know about the population model, making the comparison of the results with those obtained from simulations difficult.

Diyarbakirlioglu et al. (2022), the authors provide a real-world data application of the original method developed in Satman and Diyarbakirlioglu (2015). Specifically, they focus on the impact of the measurement error on the *market risk factor* for a large number of test assets across three popular asset pricing models, namely the Capital Asset Pricing Model, the Fama-French three-factor model and the Fama-French five-factor model. We thus leave the investigation of the behaviour of our method in these financial models, which are basically multivariate-multiple regressions, for future work.

Disclosure statement

We hereby certify that the material presented in the manuscript is the authors' original work, currently not being considered for publication elsewhere. The content therefore reflects the authors' own research in a truthful and complete manner. There are no conflicts of interest.

References

Ashenfelter, O. and Krueger, A. (1994). Estimates of the economic returns to schooling from a new sample of twins. *American Economic Review*, 84(5):1157–73.

Baluja, S. (1994). Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning.

Berkson, J. (1950). Are there two regressions? *Journal of the Americal Statistical Association*, 45(250):164–180.

Bound, J., Brown, C., and Mathiowetz, N. (2001). Measurement error in survey data. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 5, pages 3705–3843. Elsevier. Section: 59.

Buonaccorsi, J. P. (2010). *Measurement Error: Models, Methods, and Applications*. Chapman & Hall.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman & Hall.

Chen, X., Hong, H., and Nekipelov, D. (2011). Nonlinear models of measurement errors. *Journal of Economic Literature*, 49(4):901–937.

Cheng, C.-L. and Van Ness, J. W. (1999). *Statistical Regression with Measurement Error*. Wiley.

Cragg, J. G. (1994). Making good inferences from bad data. *Canadian Journal of Economics*, 27(4):776–800.

Dagenais, M. G. and Dagenais, D. L. (1997). Higher moment estimators for linear regression models with errors in the variables. *Journal of Econometrics*, 76(1):193–221.

Davidson, R. and MacKinnon, J. G. (2004). *Econometric Theory and Methods*. Oxford University Press.

Diyarbakirlioglu, E., Desban, M., and Lajili-Jarjir, S. (2022). Asset pricing models with measurement error problems: A new framework with compact genetic algorithms. *Finance*, 43(2):1–78.

Durbin, J. (1954). Errors in variables. *International Statistical Review*, 22(1):23–32.

Feng, Z., Zhang, J., and Chen, Q. (2020). Statistical inference for linear regression models with additive distortion measurement errors. *Statistical Papers*, 61(6):2483–2509.

Frisch, R. (1934). *Statistical confluence analysis by means of complete regression systems*. Universitetets Okonomiske Instituut.

Fuller, W. A. (1987). *Measurement Error Models*. John Wiley & Sons, Inc.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc.

Griliches, Z. (1987). Economic data issues. In Griliches, Z. and Intriligator, M. D., editors, *Handbook of Econometrics*, volume 3, chapter 25, pages 1465–1514. Elsevier, 1 edition.

Harik, G. R., Lobo, F. G., and Goldberg, D. E. (1999). The compact genetic algorithm. *IEEE Transactions on Evolutionary Computation*, 3(4):287–297.

Harik, G. R., Lobo, F. G., and Sastry, K. (2006). *Linkage Learning via Probabilistic Modeling in the Extended Compact Genetic Algorithm (ECGA)*, pages 39–61. Springer Berlin Heidelberg, Berlin, Heidelberg.

Harmon, C. and Walker, I. (1995). Estimates of the economic return to schooling for the united kingdom. *American Economic Review*, 85(5):1278–86.

Hausman, J. A. (2001). Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *Journal of Economic*

Perspectives, 15(4):57–67. Publisher: American Economic Association.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. University of Michigan Press.

Holland, J. H. (1987). Genetic algorithms and classifier systems: Foundations and future directions. In *ICGA*, pages 82–89.

Hyslop, D. R. and Imbens, G. W. (2001). Bias from classical and other forms of measurement error. *Journal of Business & Economic Statistics*, 19(4):475–481.

Klepper, S. and Leamer, E. E. (1984). Consistent sets of estimates for regressions with errors in all variables. *Econometrica*, 52(1):163–183.

Larranaga, P. (2002). A review on estimation of distribution algorithms. In *Estimation of distribution algorithms*, pages 57–100. Springer.

Pelikan, M., Goldberg, D. E., and Lobo, F. G. (2002). A survey of optimization by building and using probabilistic models. *Computational Optimization and Applications*, 21(1):5–20.

Racicot, F.-E. (2015). Erreurs de mesure sur les variables économiques et financières. *La Revue des Sciences de Gestion*, 3-4(267-268):79–103.

Satman, M. H. and Diyarbakirlioglu, E. (2015). Reducing errors-in-variables bias in linear regression using compact genetic algorithms. *Journal of Statistical Computation and Simulation*, 85(16):3216–3235.

Satman, M. H. and Diyarbakirlioglu, E. (2022). *eive: An Algorithm for Reducing Errors-in-variables Bias in Linear Regression*. R package version 3.1.0.

Shalabh, Garg, G., and Misra, N. (2010). Consistent estimation of regression coefficients in ultrastructural measurement error model using stochastic prior information. *Statistical Papers*, 51(3):717–748.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross-Section and Panel Data*. MIT Press.

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach*. South-Western Cengage Learning.

Table 1: Simulation results, $N = 2$

		$\widehat{\beta}_0^{LS}$			$\widehat{\beta}_0^{mCGA}$			$\widehat{\beta}_1^{LS}$			$\widehat{\beta}_1^{mCGA}$		
		bias	var	MSE	bias	var	MSE	bias	var	MSE	bias	var	MSE
T = 30													
σ_ϵ	σ_δ												
1.0	0.5	-0.0100	0.2129	0.2130	-0.0144	0.2717	0.2719	-1.0021	0.1736	1.1778	0.0876	0.2419	0.2495
1.0	0.9	-0.0001	0.4900	0.4900	-0.0222	0.8026	0.8031	-2.2521	0.2413	5.3134	-0.1240	0.6710	0.6863
3.0	0.5	-0.0097	0.4880	0.4881	-0.0178	0.6680	0.6683	-1.0029	0.4324	1.4381	0.8016	0.8770	1.5195
3.0	0.9	0.0324	0.7209	0.7220	0.0294	1.1841	1.1850	-2.2389	0.4123	5.4251	0.4999	1.1643	1.4142
T = 50													
σ_ϵ	σ_δ												
1.0	0.5	-0.0018	0.1215	0.1215	0.0019	0.1580	0.1580	-0.9963	0.1004	1.0930	0.0937	0.1704	0.1792
1.0	0.9	0.0290	0.2489	0.2498	0.0101	0.4336	0.4337	-2.2391	0.1383	5.1518	-0.0721	0.3637	0.3689
3.0	0.5	0.0186	0.3179	0.3182	0.0071	0.4276	0.4276	-0.9966	0.2343	1.2275	0.8482	0.4902	1.2096
3.0	0.9	-0.0023	0.4195	0.4195	-0.0444	0.7585	0.7604	-2.2527	0.2444	5.3189	0.6207	0.7463	1.1315
T = 100													
σ_ϵ	σ_δ												
1.0	0.5	-0.0053	0.0626	0.0626	-0.0068	0.0763	0.0764	-0.9941	0.0484	1.0366	0.0914	0.0804	0.0887
1.0	0.9	0.0028	0.1230	0.1230	-0.0146	0.2121	0.2123	-2.2369	0.0684	5.0720	0.0161	0.2227	0.2230
3.0	0.5	-0.0055	0.1580	0.1581	-0.0043	0.2005	0.2005	-0.9982	0.1096	1.1060	0.8713	0.2374	0.9965
3.0	0.9	-0.0489	0.2259	0.2283	-0.0465	0.3787	0.3809	-2.2424	0.1995	5.1480	0.7515	0.3820	0.9468

The table shows the simulation results for the model $Y_{ti} = 5 + 5X_t + \epsilon_{ti}$, where $i = 1, 2$ is the number of dependent variables and $t = 1, ..., T$ the sample size. We consider four σ_ϵ & σ_δ pairs for the regression error and the measurement error in X_t's as $X_t = X_t^\star + \delta_t$. Each configuration has been run 1000 times. Table entries show the bias, variance and MSE of the regression intercept and slope estimates from least squares and mCGA methods.

Table 2: Simulation results, $N = 5$

		$\widehat{\beta}_0^{LS}$			$\widehat{\beta}_0^{mCGA}$			$\widehat{\beta}_1^{LS}$			$\widehat{\beta}_1^{mCGA}$		
		bias	var	MSE	bias	var	MSE	bias	var	MSE	bias	var	MSE
T = 30													
σ_ϵ	σ_δ												
1.0	0.5	-0.0195	0.1971	0.1974	-0.0140	0.2486	0.2488	-0.9876	0.1713	1.1467	0.0131	0.2748	0.2750
1.0	0.9	-0.0036	0.4200	0.4200	-0.0020	0.6908	0.6908	-2.2589	0.2557	5.3582	-0.2095	0.5540	0.5979
3.0	0.5	-0.0296	0.4699	0.4708	-0.0187	0.5622	0.5626	-0.9715	0.4332	1.3770	0.3680	0.7208	0.8562
3.0	0.9	0.0044	0.6706	0.6706	-0.0058	1.0244	1.0244	-2.2260	0.3907	5.3458	0.1230	0.9173	0.9324
T = 50													
σ_ϵ	σ_δ												
1.0	0.5	-0.0131	0.1167	0.1169	-0.0066	0.1461	0.1461	-1.0052	0.1030	1.1135	0.0166	0.1650	0.1653
1.0	0.9	-0.0091	0.2602	0.2602	0.0023	0.4494	0.4494	-2.2357	0.1361	5.1346	-0.0812	0.4095	0.4161
3.0	0.5	0.0066	0.2818	0.2819	0.0048	0.3353	0.3353	-1.0225	0.2312	1.2766	0.3303	0.4015	0.5106
3.0	0.9	0.0244	0.4208	0.4214	0.0015	0.6434	0.6434	-2.2359	0.2381	5.2375	0.1811	0.6141	0.6469
T = 100													
σ_ϵ	σ_δ												
1.0	0.5	-0.0018	0.0603	0.0603	-0.0033	0.0744	0.0744	-1.0065	0.0543	1.0674	0.0229	0.0823	0.0828
1.0	0.9	-0.0051	0.1177	0.1177	-0.0147	0.2136	0.2138	-2.2379	0.0749	5.0832	-0.0501	0.1953	0.1978
3.0	0.5	-0.0072	0.1346	0.1346	-0.0079	0.1665	0.1666	-1.0177	0.1043	1.1400	0.3513	0.1835	0.3069
3.0	0.9	-0.0009	0.1982	0.1982	-0.0036	0.3197	0.3197	-2.2449	0.1185	5.1580	0.2349	0.3429	0.3981

The table shows the simulation results for the model $Y_{ti} = 5 + 5X_t + \epsilon_{ti}$, where $i = 1, ..., 5$ is the number of dependent variables and $t = 1, ..., T$ the sample size. We consider four σ_ϵ & σ_δ pairs for the regression error and the measurement error in X_t's as $X_t = X_t^\star + \delta_t$. Each configuration has been run 1000 times. Table entries show the bias, variance and MSE of the regression intercept and slope estimates from least squares and CGA methods.

Table 3: Simulation results, $N = 25$

		$\widehat{\beta}_0^{LS}$			$\widehat{\beta}_0^{mCGA}$			$\widehat{\beta}_1^{LS}$			$\widehat{\beta}_1^{mCGA}$		
		bias	var	MSE	bias	var	MSE	bias	var	MSE	bias	var	MSE
T = 30													
σ_ϵ	σ_δ												
1.0	0.5	-0.0032	0.2222	0.2222	0.0072	0.2547	0.2547	-1.0087	0.1717	1.1891	-0.0445	0.2671	0.2691
1.0	0.9	0.0071	0.4104	0.4104	-0.0080	0.6660	0.6661	-2.2353	0.2545	5.2509	-0.1845	0.6258	0.6598
3.0	0.5	-0.0154	0.5158	0.5161	-0.0117	0.5778	0.5779	-0.9725	0.4306	1.3763	0.0969	0.6313	0.6407
3.0	0.9	0.0110	0.6817	0.6818	0.0325	0.9573	0.9583	-2.2341	0.4196	5.4106	-0.1492	0.9235	0.9457
T = 50													
σ_ϵ	σ_δ												
1.0	0.5	0.0125	0.1198	0.1199	0.0119	0.1476	0.1477	-1.0225	0.1086	1.1540	0.0037	0.1592	0.1593
1.0	0.9	-0.0308	0.2465	0.2475	-0.0144	0.3983	0.3985	-2.2302	0.1392	5.1129	-0.0969	0.4075	0.4169
3.0	0.5	-0.0126	0.2707	0.2709	-0.0108	0.3081	0.3082	-1.0134	0.2248	1.2519	0.0283	0.3352	0.3360
3.0	0.9	0.0156	0.4292	0.4294	0.0074	0.6543	0.6544	-2.2545	0.2419	5.3246	-0.0456	0.6059	0.6079
T = 100													
σ_ϵ	σ_δ												
1.0	0.5	0.0083	0.0601	0.0602	0.0121	0.0730	0.0731	-0.9960	0.0500	1.0420	0.0127	0.0785	0.0787
1.0	0.9	0.0224	0.1279	0.1284	0.0263	0.2120	0.2127	-2.2544	0.0749	5.1572	-0.0433	0.2070	0.2089
3.0	0.5	0.0085	0.1469	0.1470	0.0117	0.1607	0.1609	-0.9946	0.1172	1.1065	0.0717	0.1768	0.1820
3.0	0.9	-0.0040	0.2054	0.2054	-0.0017	0.3125	0.3125	-2.2386	0.1234	5.1349	0.0053	0.2884	0.2885

Notes: The table shows the simulation results for the model $Y_{ti} = 5 + 5X_t + \epsilon_{ti}$, where $i = 1, \ldots, 25$ is the number of dependent variables and $t = 1, \ldots, T$ the sample size. We consider four σ_ϵ & σ_δ pairs for the regression error and the measurement error in X_t's as $X_t = X_t^\star + \delta_t$. Each configuration has been run 1000 times. Table entries show the bias, variance and MSE of the regression intercept and slope estimates from least squares and CGA methods.

Table 4: Simulation results for β_1

σ_ϵ	σ_δ	T	Bias $\widehat{\beta}_1^{LS}$	Bias $\widehat{\beta}_1^{mCGA}$	$Var(\widehat{\beta}_1^{LS})$	$Var(\widehat{\beta}_1^{mCGA})$	MSE $\widehat{\beta}_1^{LS}$	MSE $\widehat{\beta}_1^{mCGA}$
			Panel A: Number of dependent variables, N = 2					
1	0.5	30	-1.0021	0.0876	0.1736	0.2419	1.1778	0.2495
		50	-0.9963	0.0937	0.1004	0.1704	1.0930	0.1792
		100	-0.9941	0.0914	0.0484	0.0804	1.0366	0.0887
1	0.9	30	-2.2521	-0.1240	0.2413	0.6710	5.3134	0.6863
		50	-2.2391	-0.0721	0.1383	0.3637	5.1518	0.3689
		100	-2.2369	0.0161	0.0684	0.2227	5.0720	0.2230
3	0.5	30	-1.0029	0.8016	0.4324	0.8770	1.4381	1.5195
		50	-0.9966	0.8482	0.2343	0.4902	1.2275	1.2096
		100	-0.9982	0.8713	0.1096	0.2374	1.1060	0.9965
3	0.9	30	-2.2389	0.4999	0.4123	1.1643	5.4251	1.4142
		50	-2.2527	0.6207	0.2444	0.7463	5.3189	1.1315
		100	-2.2424	0.7515	0.1995	0.3820	5.1480	0.9468
			Panel B: N = 5					
1	0.5	30	-0.9876	0.0131	0.1713	0.2748	1.1467	0.2750
		50	-1.0052	0.0166	0.1030	0.1650	1.1135	0.1653
		100	-1.0065	0.0229	0.0543	0.0823	1.0674	0.0828
1	0.9	30	-2.2589	-0.2095	0.2557	0.5540	5.3582	0.5979
		50	-2.2357	-0.0812	0.1361	0.4095	5.1346	0.4161
		100	-2.2379	-0.0501	0.0749	0.1953	5.0832	0.1978
3	0.5	30	-0.9715	0.3680	0.4332	0.7208	1.3770	0.8562
		50	-1.0225	0.3303	0.2312	0.4015	1.2766	0.5106
		100	-1.0177	0.3513	0.1043	0.1835	1.1400	0.3069
3	0.9	30	-2.2260	0.1230	0.3907	0.9173	5.3458	0.9324
		50	-2.2359	0.1811	0.2381	0.6141	5.2375	0.6469
		100	-2.2449	0.2349	0.1185	0.3429	5.1580	0.3981
			Panel C: N = 25					
1	0.5	30	-1.0087	-0.0445	0.1717	0.2671	1.1891	0.2691
		50	-1.0225	0.0037	0.1086	0.1592	1.1540	0.1593
		100	-0.9960	0.0127	0.0500	0.0785	1.0420	0.0787
1	0.9	30	-2.2353	-0.1845	0.2545	0.6258	5.2509	0.6598
		50	-2.2302	-0.0969	0.1392	0.4075	5.1129	0.4169
		100	-2.2544	-0.0433	0.0749	0.2070	5.1572	0.2089
3	0.5	30	-0.9725	0.0969	0.4306	0.6313	1.3763	0.6407
		50	-1.0134	0.0283	0.2248	0.3352	1.2519	0.3360
		100	-0.9946	0.0717	0.1172	0.1768	1.1065	0.1820
3	0.9	30	-2.2341	-0.1492	0.4196	0.9235	5.4106	0.9457
		50	-2.2545	-0.0456	0.2419	0.6059	5.3246	0.6079
		100	-2.2386	0.0053	0.1234	0.2884	5.1349	0.2885

The table shows the bias, variance and the mean-square error of the regression slope estimates $\widehat{\beta}_1^{mCGA}$. The data generating process is specified as $Y_{ti} = 5 + 5X_t + \epsilon_{ti}$, where $i = 1, ..., N$ is the number of dependent variables and $t = 1, ..., T$ the sample size. A given configuration is described by four parameters, namely the standard deviation of regression error σ_ϵ, the standard deviation of measurement error σ_δ, the sample size T, and the number of dependent variables N.

Table 5: Simulation results for β_0

σ_ϵ	σ_δ	T	Bias $\widehat{\beta}_0^{LS}$	Bias $\widehat{\beta}_0^{mCGA}$	$Var(\widehat{\beta}_0^{LS})$	$Var(\widehat{\beta}_0^{mCGA})$	MSE $\widehat{\beta}_0^{LS}$	MSE $\widehat{\beta}_0^{mCGA}$
			Panel A: Number of dependent variables, N = 2					
1	0.5	30	-0.0100	-0.0144	0.2129	0.2717	0.2130	0.2719
		50	-0.0018	0.0019	0.1215	0.1580	0.1215	0.1580
		100	-0.0053	-0.0068	0.0626	0.0763	0.0626	0.0764
1	0.9	30	-0.0001	-0.0222	0.4900	0.8026	0.4900	0.8031
		50	0.0290	0.0101	0.2489	0.4336	0.2498	0.4337
		100	0.0028	-0.0146	0.1230	0.2121	0.1230	0.2123
3	0.5	30	-0.0097	-0.0178	0.4880	0.6680	0.4881	0.6683
		50	0.0186	0.0071	0.3179	0.4276	0.3182	0.4276
		100	-0.0055	-0.0043	0.1580	0.2005	0.1581	0.2005
3	0.9	30	0.0324	0.0294	0.7209	1.1841	0.7220	1.1850
		50	-0.0023	-0.0444	0.4195	0.7585	0.4195	0.7604
		100	-0.0489	-0.0465	0.2259	0.3787	0.2283	0.3809
			Panel B: N = 5					
1	0.5	30	-0.0195	-0.0140	0.1971	0.2486	0.1974	0.2488
		50	-0.0131	-0.0066	0.1167	0.1461	0.1169	0.1461
		100	-0.0018	-0.0033	0.0603	0.0744	0.0603	0.0744
1	0.9	30	-0.0036	-0.0020	0.4200	0.6908	0.4200	0.6908
		50	-0.0091	0.0023	0.2602	0.4494	0.2602	0.4494
		100	-0.0051	-0.0147	0.1177	0.2136	0.1177	0.2138
3	0.5	30	-0.0296	-0.0187	0.4699	0.5622	0.4708	0.5626
		50	0.0066	0.0048	0.2818	0.3353	0.2819	0.3353
		100	-0.0072	-0.0079	0.1346	0.1665	0.1346	0.1666
3	0.9	30	0.0044	-0.0058	0.6706	1.0244	0.6706	1.0244
		50	0.0244	0.0015	0.4208	0.6434	0.4214	0.6434
		100	-0.0009	-0.0036	0.1982	0.3197	0.1982	0.3197
			Panel C: N = 25					
1	0.5	30	-0.0032	0.0072	0.2222	0.2547	0.2222	0.2547
		50	0.0125	0.0119	0.1198	0.1476	0.1199	0.1477
		100	0.0083	0.0121	0.0601	0.0730	0.0602	0.0731
1	0.9	30	0.0071	-0.0080	0.4104	0.6660	0.4104	0.6661
		50	-0.0308	-0.0144	0.2465	0.3983	0.2475	0.3985
		100	0.0224	0.0263	0.1279	0.2120	0.1284	0.2127
3	0.5	30	-0.0154	-0.0117	0.5158	0.5778	0.5161	0.5779
		50	-0.0126	-0.0108	0.2707	0.3081	0.2709	0.3082
		100	0.0085	0.0117	0.1469	0.1607	0.1470	0.1609
3	0.9	30	0.0110	0.0325	0.6817	0.9573	0.6818	0.9583
		50	0.0156	0.0074	0.4292	0.6543	0.4294	0.6544
		100	-0.0040	-0.0017	0.2054	0.3125	0.2054	0.3125

The table shows the bias, variance and the mean-square error of the regression intercept estimates $\widehat{\beta}_0^{mCGA}$. The data generating process is specified as $Y_{ti} = 5 + 5X_t + \epsilon_{ti}$, where $i = 1, ..., N$ is the number of dependent variables and $t = 1, ..., T$ the sample size. A given configuration is described by four parameters, namely the standard deviation of regression error σ_ϵ, the standard deviation of measurement error σ_δ, the sample size T, and the number of dependent variables N.

Table 6: Artificial multivariate errors-in-variables model dataset

t	Y_1	Y_2	$X^\star$	δ	X	W_1	W_2
1	7.882	7.940	-0.560	-0.843	-1.404	0.253	1.026
2	2.539	2.518	-0.230	0.419	0.189	-0.029	-0.285
3	6.229	6.554	1.559	0.077	1.635	-0.043	-1.221
4	12.755	12.140	0.071	-0.569	-0.499	1.369	0.181
5	2.872	3.752	0.129	0.627	0.756	-0.226	-0.139
6	21.141	22.631	1.715	0.213	1.928	1.516	0.006
7	0.702	1.939	0.461	-0.148	0.313	-1.549	0.385
8	-1.923	-0.214	-1.265	0.448	-0.817	0.585	-0.371
9	5.027	4.984	-0.687	0.439	-0.248	0.124	0.644
10	3.668	0.696	-0.446	0.411	-0.035	0.216	-0.220
11	14.102	15.809	1.224	0.344	1.568	0.380	0.332
12	10.380	8.311	0.360	0.277	0.637	-0.502	1.097
13	5.896	8.254	0.401	-0.031	0.370	-0.333	0.435
14	-1.225	0.740	0.111	-0.153	-0.042	-1.019	-0.326
15	3.125	1.162	-0.556	-0.190	-0.746	-1.072	1.149
16	20.721	21.122	1.787	-0.347	1.440	0.304	0.994
17	12.578	12.210	0.498	-0.104	0.394	0.448	0.548
18	-4.015	-4.947	-1.967	-0.633	-2.599	0.053	0.239
19	9.129	8.464	0.701	1.084	1.786	0.922	-0.628
20	18.666	18.088	-0.473	0.604	0.131	2.050	1.361
21	-5.678	-6.326	-1.068	-0.562	-1.629	-0.491	-0.600
22	2.353	1.839	-0.218	-0.201	-0.419	-2.309	2.187
23	12.071	13.250	-1.026	-0.233	-1.259	1.006	1.533
24	-3.625	-1.269	-0.729	0.390	-0.339	-0.709	-0.236
25	-4.853	-7.984	-0.625	-0.042	-0.667	-0.688	-1.026

The table shows the dataset used in the empirical example. Y_1 and Y_2 are the response variables. $X^\star$ refers to the true regressor, with no measurement error. X is the error-prone regressor, defined as $X = X^\star + \delta$. W_1 and W_2 are additional regressors, generated with no errors-in-variables.

Table 7: Estimation results

	Model 1: OLS	Model 2: EIV	Model 3: mCGA
β	5.048	3.931	4.662
	(0.174)	(0.418)	(0.208)
γ_1	4.86	4.592	5.489
	(0.169)	(0.481)	(0.215)
γ_2	4.776	5.199	5.118
	(0.197)	(0.561)	(0.255)
α	4.789	4.434	4.436
	(0.169)	(0.476)	(0.217)
Observations	25	25	25
Adjusted R^2	0.989	0.918	0.982

The table shows the estimation results for the regression model $Y_{ti} = \alpha + \beta X_t^\star + \gamma_1 W_{t1} + \gamma_2 W_{t2} + \epsilon_{ti}$ using the artificial dataset for the first response variable Y_1. The population parameters are all equal to 5. Model 1 refers to the initial case where the true observations on $X^\star$ are assumed to be available. Model 2 considers the classical errors-in-variables case where the values $X^\star$ are observed with error as $X = X^\star + \delta$. Model 3 shows the output obtained using the mCGA method.

JAME

JOURNAL OF APPLIED MICROECONOMETRICS

HOLISTENCE
publications