WWW.JAQM.RO

JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods Inquires

Vol. 14
No. 1
Spring
2019

ISSN 1842–4562

JAQM Editorial Board

JAQM Advisory Board

Contents

SYSTEMIC RISK OF NON PERFORMING LOANS MARKET. THE ITALIAN CASE[1]

Anna Maria D'ARCANGELIS
Department of Economics and Business,
Universitaà degli Studi della Tuscia, Viterbo, Italy

E-mail: annamaria.darcangelis@gmail.com

Giulia ROTUNDO
Department of Statistical Sciences,
Sapienza University of Rome, Rome, Italy

E-mail: giulia.rotundo@live.com

Abstract

The paper considers the systemic risk due to the Non-Performing Loans in the balance sheets of banks. Using empirical data on Non-Performing Loans of Italian Banks and following the proposal of securitization of problematic loans, we propose the use of a bipartite network to simulate a hypothetical market of asset classes and investors. A default cascade dynamic runs when asset classes are hit by multiple shocks and propagation increases losses faced by investors through both direct and indirect exposure. Our results show that the degree of differentiation of the market, with a parameter that controls the sensitivity to losses of investors, is crucial to determine the systemic risk of this kind of market.

Key words: Overlapping portfolios; Systemic Risk; Non-Performing-Loans

1. Introduction

The 2007-09 financial crisis, originated in the US credit system, overwhelmed intermediaries and markets all over the world leading to a sharp drop in both financial and economic activity. One of the main outcomes of the crisis was the propagation of shocks across countries and sectors, so as to depict it as the worst financial meltdown since the worldwide economic depression that took place during the 1930s. Subsequent to the onset of the financial crisis, the turmoil developed into liquidity and solvency crises, involving real economy. Narrowing the focus on the commercial banking system, a significant number of European banks were burdened with a high share of Non-Performing Loans (NPLs) to total gross loans: loans to public administrations, financial and non-financial firms, families for consumption, mortgages, and other types of loans suffered the economic stagnation and became problematic, revealing their financial weight on bank's balance sheets.

The buildup of this problematic amount of NPLs on the balance sheets of banks is a common feature of many countries: NPLs stood at about 1 trillion in the European Union (over 9 percent of EU GDP) at the end of 2014, more than double the 2009 level (Aiyar *et al.*, 2015). A recent keynote speech by Constancio (Constancio, 2017) asserts that "NPLs are a problem with a clear European dimension, as even those countries where banks do not struggle with asset quality, are likely to be affected by spillovers, both financial and real". Nevertheless, NPL levels rest a major concern in the southern part of the euro area, as well as in several Eastern and Southeastern European countries. This outcome was the result of several related issues: although the widespread increase of credit risk undertaken, it is evident that in some countries like Italy (and other peripheral economies) there were other causes that affected the market. Many structural causes contributed to the protracted recession phase: for instance inadequate incentives for financial institutions to write off or to sell problematic loans, low provisioning that create pricing gaps between book valuations and market values of loans and wide bid-ask spreads, the reliance on collaterals -which compress the incentives to sell-, and tax disincentives to provisioning and write-offs. From the demand side, it is worth to signal the lengthy and inefficient judicial process (more than seven years to complete a bankruptcy procedure and three years to foreclose on real estate collateral) and the lack of equity capital. All the causes have negatively influenced the demand for bad loans and the subsequent profitability and efficiency of the market (Kang & Jassaud, 2015). In this framework, the impact of negative interest rates, due to monetary policies aimed at reducing other problems, could somehow worsen the situation. This combination of causes produced a strong obstacle towards the efficient disposal of NPLs compared with the rest of developed countries. In a first stage of the crisis, Italian financial institutions were not unscathed by the turmoil, to the point that no public funds had to be injected to sustain financial institutions. However, after some years the consequences of the crisis became evident. In detail, losses due to NPLs impact on banks' profits, deteriorate capital structure determining lack of capital and drives up the regulatory capital charge, limiting the credit provision to the firms: the effect on firms' performances and on whole economy is therefore negative.

The main countermeasures proposed by the Italian State had the specific target to help the banks to get rid of bad loans, so to break the credit crunch and to sustain the economic growth. One of these proposals is the constitution of a distressed securities closed end fund, designed to invests or co-invests in securitization structures that can quicken the disposal of NPLs, favoring the creation of an efficient market for bad loans. The NPLs-market framework is driven by a securitization process where a SPV (special purpose vehicle) creates Asset-Backed-Securities (ABS) sold in the market to finance the purchasing of NPLs from the banks. The intrinsic and implicit interconnections among agents in the market, due to the securitization process that overlaps asset portfolios to build ABS, are manifestly clear. Market operators discount the whole creditworthiness evaluation process, so a shock on borrowers underlying the securities causes an increasing detected riskiness of the securities itself. Funds operating in this market do not know the specific firms and people underlying these credits and their creditworthiness. Investors trust in securitization because pricing the worst NPL gives the lower bound of the other securities price in the market.

The aim of the paper is to highlight the possibility of growth in the systemic risk of a hypothetical Italian NPLs market due to the mechanisms planned to build such market, as a leading factor for the optimal market configuration and pricing process of securities. Our proposal is a model, based on a bipartite network model of asset classes and funds, in which we build the indirect interconnections between asset classes via correlation matrices. We statistically construct direct interconnections between asset classes and funds through the Bipartite Configuration Model (BiCM) (Squartini, 2016), assuming that if there is an increasing of NPLs for a specific sector or area the value of ABS on NPLs decrease.

Bipartite networks have been mostly used in financial literature for the representation of networks of banks and assets and the consequent dynamics of overlapping portfolios. Works like Caccioli *et al.* (2014) have developed asymptotic models to study the riskiness of a distress spreading dynamics on a bipartite network, reaching the conclusion that such networks suffer the so-called "robust yet fragile" behavior. Connectivity and leverage are the main features that drive the system from a stable to an unstable region. Other works such Huang *et al.* (2013) instead, focus on empirical Bank-Assets networks, trying to reproduce the bankruptcy dynamics of recent financial crises (2007-2008) in order to develop a new systemic risk-detecting framework. Others like Miranda (2013) apply systemic risk models to identify the riskiness of different bipartite empirical networks like Brazilian Firm-Bank networks.

These researches highlight a crucial role of banks in triggering distress through financial networks. Our hypothesis is, otherwise, different. In fact, banks do not belong to our model, as the transfer of problematic loans to the funds removes them from any active role in the model and its simulations[2]. The ultimate aim is to identify the way to detect systemic risk in such bipartite empirical stylized NPLs market of Funds and asset classes. The research has been inspired by the proposal of Quaestio Capital Management SGR S.p.A., a financial advisory firm appointed to create and manage the Atlante Fund with the aim to create an NPLs Italian market.

2. Data and Model assumptions

This section describes the data and the models used for the analysis of the propagation of stress due to the non-performing loans (NPLs) of a sample of 25 Italian banks[3] that joined the closed fund Atlante[4] in April 2015.

Our analysis is based on an extensive dataset at the bank-firm level. Data on gross and net "bad loans" for individual Banks were manually collected from the notes in the Annual Report of each institution for the period 2008-2015[5]. The list of 19 Investors in the securitized NPLs has been retrieved gathering information from C&W Loan Sales 2014-2016, Debtwire, CNBC, Apax, Deloitte NPL Outlook 2014-2015, Italy24, KPMG Loan Sales Feb 2016, Quaestio.

The gross amount of the NPLs of such banks sums up to roughly 164.5 bn euros: this amount corresponds to a net value of 68 bn euros[6].

A high amount of NPLs ties up institutional capital of banks and prevents its use in the economic banking process (first and foremost lending), with negative effects on banks funding costs, profitability and credit supply. Disposal of NPLs in Italy is too slow compared with the rest of Europe due to a series of structural factors that determine an abnormal bid/ask spread:

1. Data quality, due to the lack of an organized and complete set of data in the banks databases.

2. Servicing, due to market fragmentation, lack of critical mass, critical issues affecting recovery procedures.
3. Time to recovery, a huge uncertainty in the time of recovery due to a wide divergence among the various procedures and courts.
4. Prospective trends of the Italian economy.

Banks should therefore accept a hard discount in the selling process of NPLs to specialized investors. Our hypothesis is that banks accept to liquidate the totality of net NPLs at a discount of 27.5 percent: the value of net NPLs introduced on the market would result in 49.3 bn euros, amount that would be securitized in three tranches: senior, mezzanine, and junior. This last tranche is supposed equal to 13.6 bn euros, the 27.5 percent of the selling value of NPLs. the loss suffered by the banks is ≈ 18.3 bn euros.

We can now start our process, building a bipartite model of market expositions of funds on the junior tranches of ABS.

2.1. Main data and variable description

- a vector $a \in R^{85}$ of 85 NPLs asset classes identified listing the 2015 exposure on NPLs of the 25 banks for each of the 6 customer economic activity: General Government, Other public entities, Financial companies, Insurance companies, Not Financial companies, Other entities[7].
- the NPLs sample described above was extended to the years 2008-2015. The resulting time series was used to calculate a correlation matrix $C^6 \in R^{6\times 6}$ between sectors of origins (see Appendix 3 for details) and also,
- a block-diagonal correlation matrix $C^{85} \in R^{85\times 85}$ between assets belonging to the same sector. We agree on the fact that the time series is short, but there were no data available before 2008.
- to the information of each element in vector a we add the membership of each asset to a regional area (North East (1), North West (2), Center (3), South & Islands (4)), that depends on the bank originator. Such data are reported in a vector $r \in R^{85}$. Each component has a value in the set {1, 2, 3, 4}.
- based on Quaestio (2016), we assume to deal with a set of 19 Investors. Their exposure to the NPLs is not uniform. Such peculiar number was deduced from the information reported in Quaestio (2016). Table 7 in the Appendix 3 shows the list of funds and exposures, that is the monetary amount funded by each of the external investors. The vector $f^1 \in R^{19}$ contains the numbers in the Table, that were calculated dividing the total exposure 13.6bn by the percentage of the exposure of the funds as reported in Quaestio (2016).
- through f^1, we build the adiacency exposition matrix $E \in R^{19\times 85}$. Each element E_{ij} is the amount of money that each investor i invests in any asset class j. To do this we use the Bipartite Configuration Model (BiCM) (Saracco *et al.*, 2015); Squartini (2016)). The Appendix 1-2 contain details on the BiCM.
- Network building. Our NPLs market can be represented as a directed weighted bipartite network, in which links correspond to exposures of Investors on asset classes. With this method we can construct the network of exposures for each $z \in (0, \infty]$, that is the density parameter of links. The BiCM belongs to a series of entropy-based models for bipartite networks. In a nutshell, the method estimates Investors' individual exposures as:

$$E_{ij} = \frac{z^{-1} + f_i^1 a_j}{C} a_{ij}, \quad a_{ij} = \begin{cases} 1 & with\ probability\ p_{ij} = (f_i^1\ a_j)/(z^{-1} + f_i^1 a_j) \\ 0 & otherwise \end{cases} \quad (1)$$

where f_i^1 is the amount that the Investor *i* invests on market (the element *i* of vector ***f***), a_j is the value of asset class *j*, $C = \sum_i f_i = \sum_j a_j$ e size of the market and ***Eij*** controls the presence of the link between Investor *i* and asset class *j* relted to p_{ij} that is the probability that there is a link between node *i* and *j*.

Based on the above data, we outline the ideas underlying the dynamics.

Model dynamics

The idea underling the dynamics is that the decrease of price of an asset *j* influences the system in three different ways:

a. direct exposure. Any investor *i* that is exposed on it loses a percentage of its value. Since the exposures are gathered in the matrix *E*, let us name the decrease $\Delta E\ E_{ij}$. The effect of this decrease of value does not remain confined to the funds that are directly exposed on the asset, but propagates through two other different channels;

b. indirect exposure, through the economic category. Asset classes *l* that belong to the same economic categories and that are correlated with a_j gets a decrease proportional to the correlation matrix C[85] and to the value of the exposition E_{lj}. As a consequence, the values of funds exposed to each asset class *j* decreases by $\Delta E\ E_{ij}$; *l,j*

c. indirect exposure, through the asset classes. Recalling that vector *a* reports the membership of each asset to a regional area, the asset classes *j* that are in the same region of a_i (although belonging to different economic categories) decrease their value. Again, as a consequence, the values of funds exposed to each asset class *j* decreases by $\Delta A E_{ij}$.

Let *L* be the list of assets which values decreased in [c]. The steps [a], [b], [c] are run again starting from an asset in *L* picked up randomly. Therefore, there is a sequence of decrease of values. If an investor overtakes the percentage losses threshold *q*, that we put equal for each fund for simplicity, it leaves the market and liquidates its portfolio, so all the assets in its portfolio experience a further decrease in their value. The procedure stops when there are no more investors that leave the market. When the procedure ends, we calculate the market impact of this portfolio's configurations of the market, through BiCM, revealing the amount of residual asset classes market value survived. We run the program for different configuration of the network, obtained through BiCM. The dynamic is run again with several value of the factor z, which gives the density of links in a wide range (0%-100%). Increasing z results in an increase of the density (connectivity) of the bipartite network.

In terms of variables, the pseudo-code of the model is as follows:

- At first we construct an exposition matrix E_{ij} for a given z that controls the density of the links in the network.

- At the begin of each run of the program, at time *t*, we pick up randomly an asset class *l* to shock. The three means for the propagation of the shock are represented as follows:

1. The shock has the effect to decrease of a percentage discount factor *v* the value of the asset, so any funds *i* exposed on *l* suffers a direct losses via exposition matrix E_{ij} :

$$E_{il}(t+1) = E_{il}(t) - v \cdot E_{il}(t) \quad (2)$$

2. then stress propagates among other asset classes of the same economic sector $\{j \neq l, j \in K\}$ via correlation matrix C_{85}, so we have:

$$\forall_j \neq l, j \in K,\ E_{ij}(t+1) = E_{ij}(t) - v \cdot C^{85}{}_{jl} E_{ij}(t) \quad (3)$$

Reminding that the diagonal elements of the correlation matrix are equal to one, the two steps above can be gathered into

$$\forall_j \in K,\ E_{ij}(t+1) = E_{ij}(t) - \Delta_E E_{ij} \quad (4)$$

Where

$$\forall_j \in K,\ \Delta_E E_{ij} = v \cdot C^{85}{}_{jl} \cdot E_{ij}(t)$$

3. in the third step the asset class *j* that belongs to another economic sector but comes from the same region category *R* of *l* receives a shock:

$$\forall_j = R,\ E_{ij}(t+2) = E_{ij}(t+1) - \Delta_A E_{ij}, \quad (5)$$

where

$$\Delta_A E_{ij} = v \cdot C^{6}{}_{jl} \cdot E_{ij}(t+1)$$

- In this dynamics scenario Investor *i* suffers of incremental monetary losses. If *l*, one of the Funds, has reached the escape threshold *q* at time t^*, $\frac{\sum_l E_{lj}(t^*)}{\sum_l E_{lj}(\mathrm{t})} < q$, fund *l* sells its assets that depreciate:

$$E_{lj}(t^*+1) = E_{lj}(t^*) \cdot \left(1 - \frac{\sum_{ij} E_{ij}(t^*)}{\sum_{ij} E_{ij}(t)}\right) \quad (6)$$

- If in the previous step no one has reached the escape threshold, the dynamics stops, otherwise it restarts with an equal exogenous shock to another asset class $j \in R$ picked at random.
- When the dynamics stops in $\hat{t}$ we calculate the market impact of the cascading process in terms of monetary loss in the market at any step monitoring the devaluation of fund's exposure plus the effect of depreciation

$$AV(z) = \frac{\sum_{ij} E_{ij}(\hat{t})}{\sum_{ij} E_{ij}(\mathrm{t})} \quad (7)$$

This is the assets value at risk of default and the impact that Funds could suffer.

3. Results

The main goal of our work is to assess the systemic impact of a dynamics on NPLs default cascade on the empirical reconstruction of a hypothetic NPLs market. To do this, using BiCM, we are able to define different random overlapping portfolios configurations in the market, just varying z, the density control parameter. In financial terms, the variation of z implies the rise of market integration or, from the Investors point of view, the increase of portfolio differentiation. Given stable financial resources, it implies also the increase of the number of exposures on the market with a lower value, going from a sparse specialized market to an integrated highly differentiated one. So the maximal density of edges brings to a situation where every Investor is exposed on the same asset class: it means that everyone purchases an amount of the market portfolio. In a network perspective we have a star in the monopartite projection of the bipartite market network. In short, it is the same situation of a Fund that purchases the whole quantity of asset classes, in line with the original aim of Atlante Fund, the Investment of the whole junior tranche of the Italian NPLs securitization.

We now present the results of our model's simulations. We run 100 simulations for a discrete series of z within its range, and we vary the control parameter of the market escape $q \in \{20\%, 50\%, 80\%\}$. First, we focus on the more realistic Investors configuration reconstructing the expositions matrix E^1_{ij} with vector F^1. Looking at the dynamics of a single realization of the system, fig.1 shows the evolution of the remaining market values or conversely the percentage impact of the default cascade. The different trend trajectories refer to different model configurations on q, (figs. 1, 2, 3), in which we select randomly an asset with the exogenous shock v = 10% of initial value. The first striking observation is that the dynamics converges to a threshold around 20%. We remark that for a low connectivity (i.e. low network density of edges or low portfolio's differentiation) the value of all the assets goes to zero. Conversely, with an increasing density we recover partial value of the market. A significant assumption is the "market escape". We define a targeting clearance threshold q to reproduce the role that in usual bipartite financial networks belongs to banks. With their balance sheet constraints, at the default time, they are candidates for intensifying the stress within the network Caccioli *et al.* (2014). In our model we define q as the losses-targeting policy constraint that Investors have and that triggers portfolio liquidation: this is in line with "panic selling" situations in the markets.

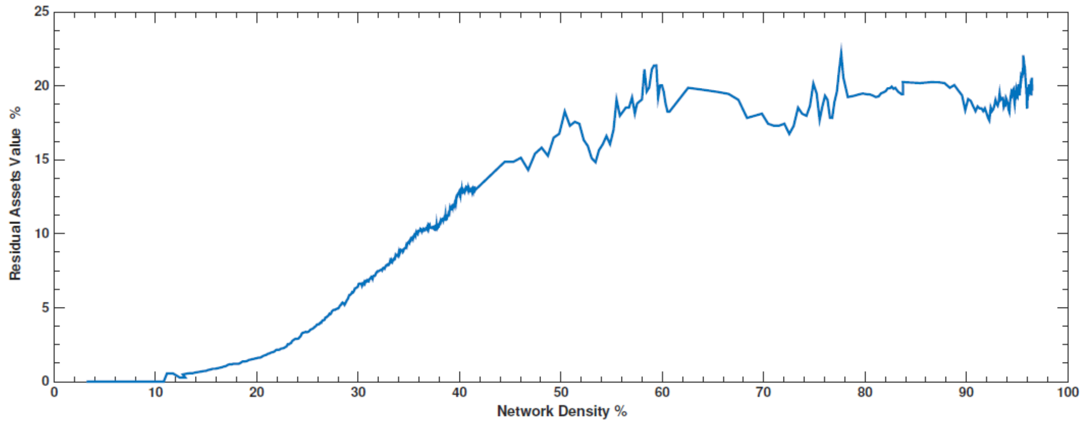


Figure 1. Final residual assets value of F^1 in the system for $q = 20\%$

The analysis of the dynamics among the various "escape-market threshold" configurations, reveals that there is convergence of the system although there are many fluctuations. The system converges to a flat threshold of $\approx (42 - 48\%)$, where we have no more differences in systemic impact of connectivity and parameter q. Increasing the residual asset classes value, the fluctuation of the result increases. The same point emerges in fig. 4, which shows a similar dynamics of "market escape". Considering the frequencies of the dynamic's "breaks", the detection of a break indicates that there are no more "escapes", with losses at a minimal level. The dynamics of "breaks" displays exactly the same evolution of residual asset classes value because they indicate which configuration of the system is gradually more resilient to the shocks. So q is one of the main parameters that lead to different trajectories.

It is worth to note that these results, based on empirical data, do not confirm the thesis which states that networks with medium connectivity are more resilient to default cascade - Caccioli *et al.* (2014). The characteristics of the network (two types of nodes and two types of links) seem to be a key factor in the present analysis. Namely, the correlation matrices induce spreading of distress beyond directed expositions. So this effect of multiplicative shocks, given by the multiple cycles in our framework design, is high when we have sparse networks. With a lower node degree and higher single exposure the effect of multiplicative shocks is higher. The conclusion with a model that reconstructs links redistributing proportionally their weights, is that low density networks with hidden connections are more fragile.

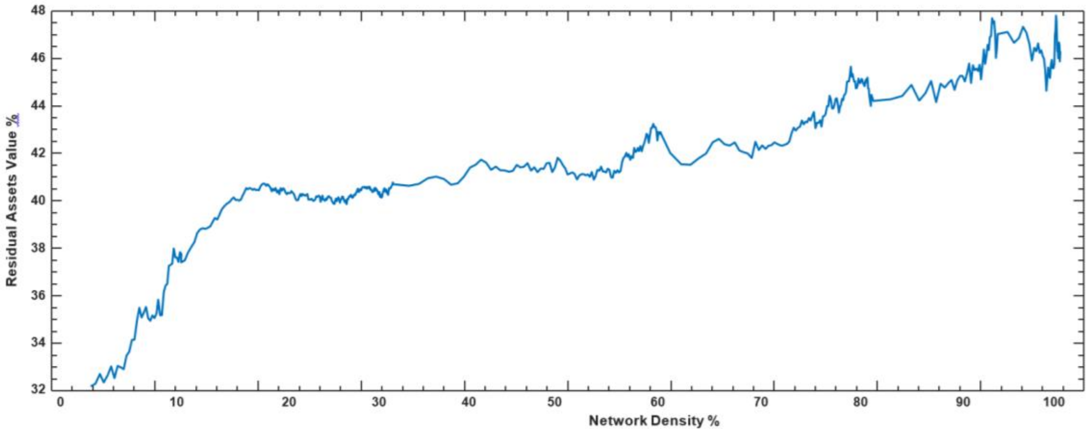


Figure 2. Final residual assets value of F^1 in the system for $q = 50\%$

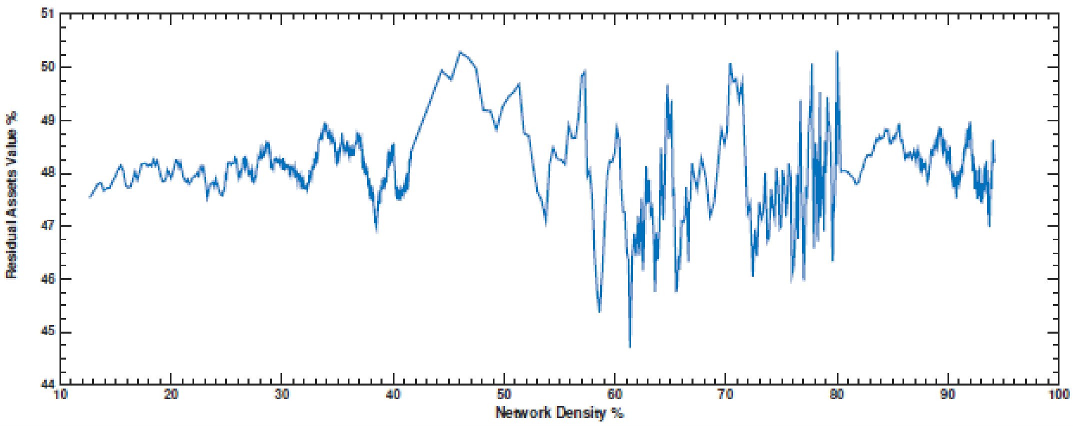


Figure 3. Final residual assets value of F^1 in the system for $q = 80\%$

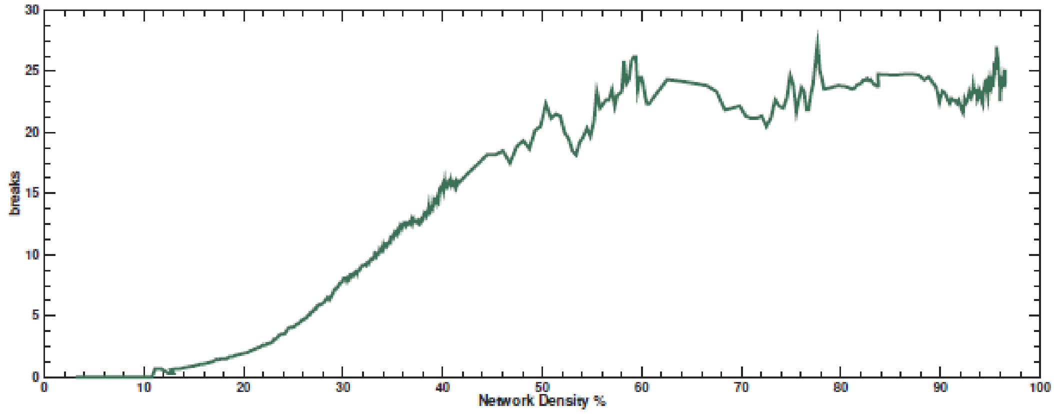


Figure 4. Frequencies of breaks of F^1 for $q = 20\%$

4. Second setting

We proceed with a different assumption on the Funds that buy the NPLs: a second hypothesis considers a large amount of investors, that invest all the same amount. Since we have 85 assets, we assumed to have 85 funds. Each component of the vector $F^2 \in R^{85}$ is equal to 159964.78. This number was calculated as a raw division 13.6bn/85=159964.78m. Of course, any other large number could have been used for the present analysis. Figs. 5, 6, 7 show the results of the dynamics when F^2 is used in order to build an exposition matrix E_{ij}^2 (85 x 85) with 85 Investors of the same dimension.

We see that there are no remarkable differences between these two configurations for both residual asset classes value and breaks. It means that the topology of the network plays the major role in the dynamics. The connectedness is, as it is well known, the main spreading channel of stress through the market.

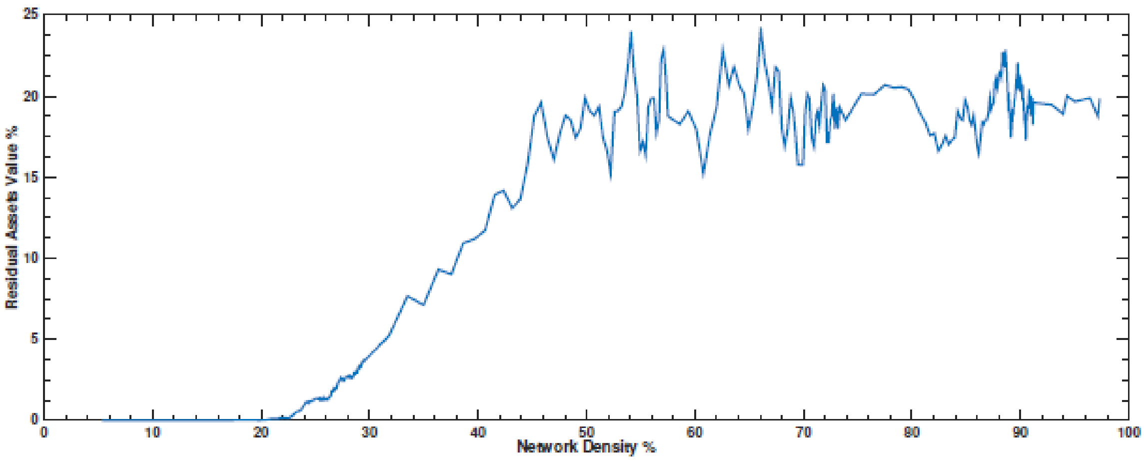


Figure 5. Final residual assets value of F^2 for $q = 20\%$

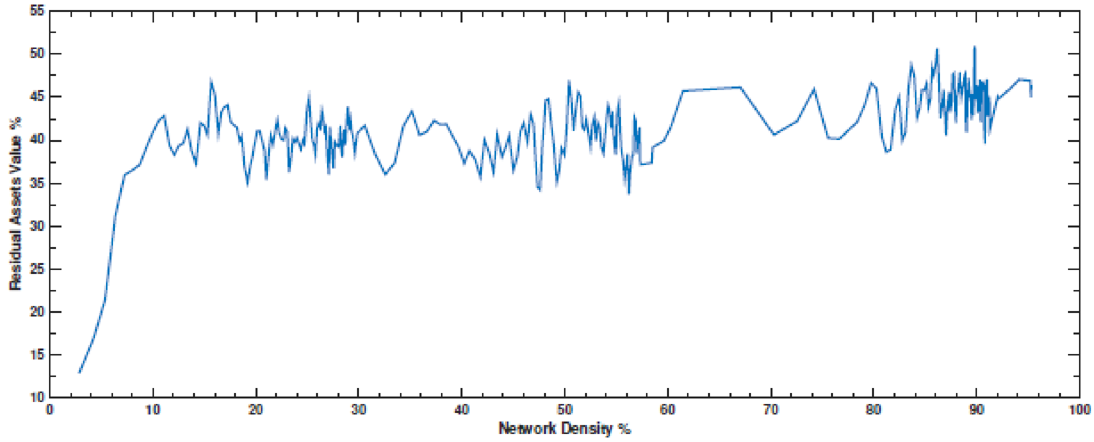


Figure 6. Final residual assets value of $F^2$for $q = 50\%$

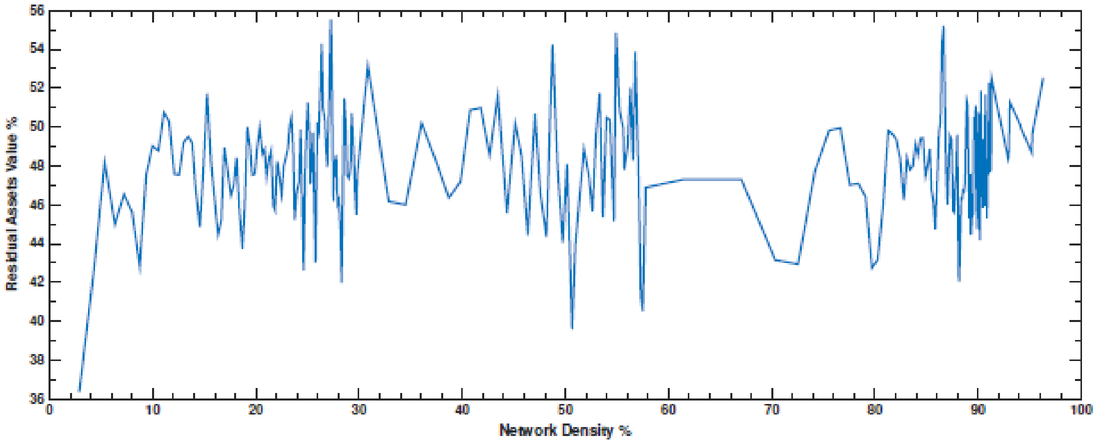


Figure 7. Final residual assets value of F^2 for $q = 80\%$

5. Conclusions

In our work we have designed an innovative empirical model to mimic the dynamics of a cascade defaults of assets in a hypothetical NPL market. The model relies on Italian banks NPLs data source coming from banks balance sheets, and is built on simple assumptions on Funds strategic behavior during periods of financial distress on portfolios assets. What emerges is that the main leading factors of stress are: the "market escape" dependence of paths, the fragility of a system with too many similar units where there are multiple connections of different nature and not relevant differences between systems of actors of various wealth. These features of NPL market are of the main relevance w.r.t. the capability of defining and quantifying the Systemic Risk of this market. We have focused on the dynamics of NPLs because of their crucial role due to the fact that they are becoming a major concern at policy level and about the correct monetary policy transmission and so for the economic growth. The results show that the control parameters of the connectivity and "market escape", v and q, lead the dynamics path. For instance, v could be managed with a scheme like Atlante where a Fund purchases all the asset classes, because it is the same situation of many funds that purchase a fraction of all the asset classes and where the bipartite network would have the maximum density.

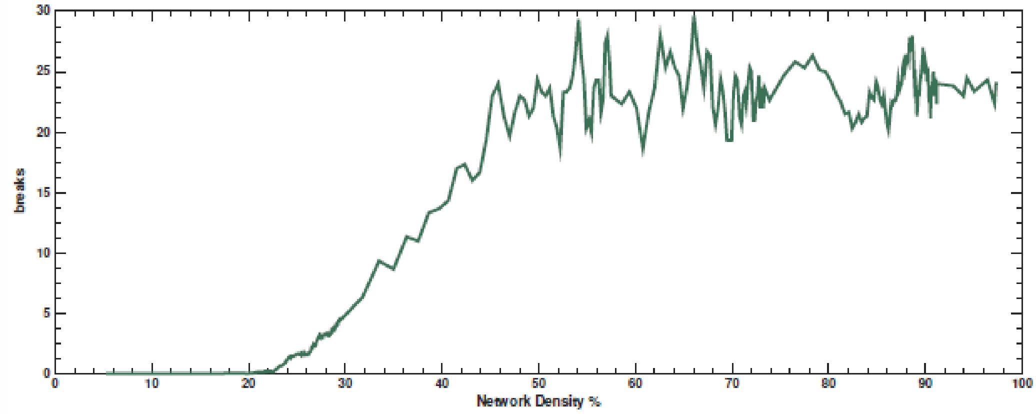


Figure 8. Frequencies of breaks of F^2 for q = 20%

Otherwise, policy makers could regulate a parameter of differentiation avoiding concentration of few portfolios on the same asset class. The parameter q could be managed by policy makers and calibrated in view of the institutions that would purchase the assets. For example, there are investors, like some Hedge funds with a short term investment duration policy, that are more sensitive to immediate losses. Investors with a lower q would exacerbate a potential fire-sale. Instead other investors, for instance Pension funds, that have a long term investment horizon, would be less sensitive to sudden losses and do not accelerate the fire-sale betting in a recovery of the investment. Therefore, policy makers could define some mechanisms to select investors and other ones to freeze the market avoiding panic-selling. This could be of interest for Policy makers, Authority and operators involved in the market optimal design and interested in evaluate the systemic impact of NPLs default cascades.

References

1. Aiyar, S., Bergthaler, W., Garrido, J. M., Ilyina, A., Jobst, A., Kang, K., Kovtun, D., Liu, Y., Monaghan, D. and Moretti, M. **A Strategy for Resolving Europe's Problem Loans,** IMF Staff Discussion Notes, 2015, 15/19.
2. Caccioli, F., Shrestha, M., Moore, C. and Farmer, J.D. **Stability analysis of financial contagion due to overlapping portfolios,** Journal of Banking & Finance, Vol. 46, 2014, pp. 233-245.
3. Cimini, R. **Il sistema di bilancio degli enti finanziari e creditizi,** Wolters Kluwer, 2016.
4. Cimini, G., Squartini, T., Gabrielli, A. and Garlaschelli, D. **Estimating Topological Properties of Weighted Networks from Limited Information,** Physical Review E, Vol. 92, Issue, 4, 2015, pp. 040802.
5. Cimini, G., Squartini, T., Garlaschelli, D. and Gabrielli, A. **Systemic Risk Analysis on Reconstructed Economic and Financial Networks,** Scientific Reports, Vol. 5, 2015, pp. 15758.
6. Constancio, V. **Resolving Europes NPL burden: challenges and benefits.** In "Tackling Europe's non-performing loans crisis: restructuring debt, reviving growth", Bruegel event, 2017.

7. Huang, X., Vodenska, I., Havlin, S. and Stanley, H.E. **Cascading Failures in Bi-partite Graphs: Model for Systemic Risk Propagation,** Nature: Scientific Reports, Vol. 3, 2013, pp. 1219.
8. Kang, H.K. & Jassaud, N. **A Strategy for Developing a Market for Non performing Loans in Italy,** Working Paper IMF, Vol. 2, 2015, 15/24.
9. De Castro Miranda, R. C. and Tabak, B. M. **Contagion Risk within Firm-Bank Bivariate Networks,** Banco Central do Brasil: working papers series, Vol. 322, 2013, pp. 1-60.
10. Park, J. & Newman, M.E.J. **Statistical mechanics of networks,** Physical Review E, Vol. 70, 2004.
11. Saracco, F., Di Clemente, R., Gabrielli, A. and Squartini, T. **Randomizing bipartite networks: the case of the World Trade Web,** Nature: Scientific Reports, Vol. 5, 2015, pp. 10595
12. Shannon, C.E. **A Mathematical Theory of Communication**, Bell System Technical Journal, Vol. 27, 1948, pp. 379-423, 623-656.
13. Squartini, T., Caldarelli, G. and Cimini, G. **Stock markets reconstruction via entropy maximization driven by fitness and density,** arXiv:1606.07684v1 [q-fin.RM], 2016.
14. Vicari, R. & Berselli, E. **La normativa di riferimento e gli schemi del bilancio bancario,** In: "Il bilancio della banca e degli altri intermediari finanziari", Egea, 2016.
15. * * * Capital Management Quaestio, Atlante Fund Presentation, 2016, *http://www .quaestiocapital.com/sites/default/files/Quaestio Atlante Presentation294 2016EN.pdf* .

[1]**Acknowledgement**
AMD and GR thank Matteo Serri for the fruitful collaboration. Furthermore, thanks to Giulio Cimini for his helpful guidances and suggestions.

[2]This implies that banks play no active role anymore in the system after the divestiture of NPLs.

[3] Intesa Sanpaolo, Unicredit, UBI Banca, Banca Popolare dell'Emilia Romagna, Banca Popolare di Milano, Credito Valtelli- nese, Banca Mediolanum, Banca Popolare di Sondrio, Banco Popolare, MPS, Iccrea Banca, Popolare di Bari, Carige, Banca Sella, CR Asti, CR Bolzano, Popolare di Puglia e Basilicata, Banco Desio, Banca di Piacenza, Banca Valsabbina, Popolare Alto Adige, Popolare Pugliese, Banca di Credito Popolare, CR Ravenna, Popolare di Cividale.

[4] Atlante is a closed-end alternative investment fund created in April 2016 by Quaestio Capital Management SGR S.p.A. The fund, regulated by Italian law and reserved for professional investors such as banks, insurance companies, banking foundations and the Cassa Depositi e Prestiti, invests at least 30% of its funds in Non-Performing Loans (NPLs) from several Italian banks. The purpose of the Fund was to promote the creation and development of a large and efficient secondary market of distressed assets in Italy.

[5] Bank of Italy classified non-performing loans in four categories: a) Bad loans ("sofferenze"), exposure to any borrower in a position of insolvency (even if insolvency is not legally ascertained) or a substantially similar situation, regardless of any loss estimate made by the bank and irrespective of any possible collateral or guarantee b) Substandard loans ("incagli"), exposure to any borrower experiencing temporary payment difficulties defined on the basis of objective factors - that the lender believes can be resolved within a reasonable period of time c) Restructured loans ("ristrutturati"), exposures in which a pool of banks or an individual bank, as a result of the deterioration of the borrowers financial situation, agree to change the original conditions (interest rate, reduction in capital, rescheduling of monthly payments, etc.), giving rise to a loss. In the event of a partial restructuring, exposure remains in its original category. d) Past due ("scaduti" or "sconfinanti"), exposure to any borrower whose loans are not included in other categories and who, at the date of the balance sheet closure have past due amounts or unauthorised overdrawn positions of more than 90 days. A new classification of loans by Bank of Italy (1/2015), transposing the technical standard published by European Banking Authority (EBA) (CE Reg. 2015/227), refers to non-performing exposures and forbearance exposures and has no impact on our sampled data. For details, see Cimini (2016).

[6] Net NPLs are defined as gross NPLs less loan loss provisions (reserves). The normative references and administrative provisions issued by the Bank of Italy (Circ.262) on the layout and presentation of Income Statement for Banks can be found in - Vicari & Berselli (2016); Cimini (2016).

[7] This classification has been extracted from the Tables related to the "Breakdown and concentration of credit exposures" in the Notes of the Annual report of each and from the Pillar III of the 25 Banks. Each bank is exposed only to some of the six customer economic activities, and the complete list has a lenght of 85.

APPENDICES

Appendix 1: Exponential Random Graph Model

To introduce the Exponential Random Graph Model (ERGM) it is useful to recall the concept of Entropy due to its fundamental role in defining Information and the way to use it to build models of networks generation.

Entropy

By definition, the amount of self-information contained in a probabilistic event depends only on the probability of that event - Shannon (1948): the smaller its probability, the larger the self-information associated with receiving the information that the event indeed occurred. Furthermore, by definition, the measure of self-information is positive and additive and the proper choice of function to quantify information, preserving this additivity, is logarithmic. If an event C is the intersection of two independent events A and B, then the amount of information at the proclamation that C has happened, equals the sum of the amounts of information at proclamations of event A and event B respectively: $I(A \cap B) = I(A) + I(B)$. Taking into account these properties, the self-information $I(wn)$ associated to the outcome *wn*with probability $P\ (wn)$ is

$$\mathrm{I}\ \ (w_n) = \log\left(\frac{1}{P\ \ (w_n)}\right) = -\log(P\ \ (w_n))$$

This definition complies with the above conditions. In the definition above, the base of the logarithm is not specified: if using base 2, the unit of $I(wn)$ is bits. Shannon defined the entropy H of a discrete random variable X with possible values $\{x1, \ldots, xn\}$ and probability mass function P(X) as:

$$H(X) = E[I(X)] = E[-\ \ln(P\ (X))]$$

Here E is the expected value operator, and I is the information content of X. I(X) is itself a random variable. When taken from a finite sample, the entropy can explicitly be written as

$$\mathrm{H(x)} = \sum_{\mathrm{i}} \mathrm{P(x_i)\ I\ (x_i)} = -\sum_{\mathrm{i}} \mathrm{P(x_i)\ \log_b P(x_i)}$$

The Exponential Random Graph Model

Now we consider a specific class of Configuration model. Consider a set G of graphs - Park & Newman (2004). Suppose to have a collection of graph observables $\{xi\}$, $i = 1 \ldots r$, that we have measured in empirical observation of some real-world networks. In practice it is often the case that we have only one measurement of an observable. In this case, however, our best estimate of the expectation value of our variable of interest is simply equal to the one measurement that we have. Let $g \in G$ be a graph in our set of graphs and let P(g) be the probability of that graph within our ensemble. We would like to choose P(g) so that the expectation value of each of our graph observables $\{xi\}$ within that distribution is equal to its observed value. The best choice of probability distribution is the one that maximizes the Gibbs-Shannon entropy

$$S = -\sum_{g=G} P(g)\ \ P(g) \ln P(g)$$

subject to the constraints

$$\sum_g P(g)\ x_i\,(g) = < x_i >$$

plus the normalization condition

$$\sum P(g) = 1$$

Here *xi*(*g*) is the value of *xi* in graph *g*. Introducing Lagrange multipliers *α*, {*ϑi*}, we then find that the maximum entropy is achieved satisfying

$$\frac{\partial}{\partial P(g)}\left[S + \alpha\left(1 - \sum_g P(g)) + \sum_i \theta_i(< x_i > - \sum_g P(g)\ x_i(g)\right)\right] = 0$$

for all graphs g this gives

$$-\ln P(g)\left(\ P(g)\ \frac{1}{P(g)}\right) - \alpha - \sum_i \theta_i\, x_i\,(g) = 0$$

or equivalent

$$e^{lnP(g)} = \quad e^{(-1-a)}\ e^{-\sum_i\ \theta_i\, x_i\,(g)}$$

where we define graph Hamiltonian H and partition function Z;

$$H(g) = \sum_i \theta_i\, x_i\,(g), \qquad Z = e^{(1+a)} = \sum_g e^{(-H(g))}$$

for normalization, and then

$$P(g) = \frac{e^{-H(g)}}{Z}$$

The last equation defines the exponential random graph model - Park & Newman (2004). The Exponential Random Graph is the distribution over a specified set of graphs that maximizes the entropy subject to the known constraints. The expected value of any graph property x within the model is simply

$$< x \quad > = \sum_g P(g)\, x(g)$$

Suppose we have the complete set {*ki*}, the degree sequence of the network. The Exponential Random Graph model appropriate to this set is the one having Hamiltonian $H = \sum_i\ \theta_i k_i$ where we now have one parameter θ_i for each vertex i. Noting that $k_i = \sum_j\ \sigma_{ij}$, this can also be written

$$H = \sum_{ij} \theta_i \sigma_{ij} \quad = \sum_{i<j} (\theta_i + \theta_j)\sigma_{ij}$$

The partition function is

$$Z = \sum_{\sigma_{ij}} \exp\left(-\sum\nolimits_{i<j} (\theta_i + \theta_j)\sigma_{ij}\right) = \Pi_{i<j}\left(\ \sum\nolimits_{\sigma_{ij}=0}^{1} e^{-(\theta_i+\theta_j\,)\sigma_{ij}}\ \right) =$$

$$= \Pi_{i<j}\left(1 + e^{-(\theta_i+\theta_j\,)}\ \right)$$

More generally, we could specify a Hamiltonian

$$H = \sum_{i<j} \Theta_{ij}\, \sigma_{ij}$$

with a separate parameter Θ*ij* coupling to each edge. Then $Z = \Pi_{i<j}\,(1 + e^{-\Theta_{ij}})$ and, defining *F* = − ln *Z*, we have

$$F = -\sum_{i<j} \ln(1 + e^{-\Theta_{ij}})$$

This allows us to calculate the probability of the occurrence *pij* of an edge between vertices *i* and *j*

$$\mathrm{p}_{ij} = < \sigma_{ij} > = \frac{1}{Z}\sum_{\mathrm{G}} \sigma_{ij} e^{-H} = -\frac{1}{Z}\frac{\partial Z}{\partial \Theta_{ij}} = \frac{\partial F}{\partial \Theta_{ij}} = \frac{e^{-\Theta_{ij}}}{1+e^{-\Theta_{ij}}}$$

Appendix 2: The configuration model

The Configuration Model (CM) (Cimini et al. (2015a,b)) investigates if it is possible to estimate topological properties of a network starting from limited information. The Fitness based Configuration model can be seen as a specific case where the set of properties {*Ci*} is the degree sequence {*ki*}, *i* = 1 . . . *n* of the nodes of the network, where the values of <*ki*> for all nodes i are fixed and each node can be identified by its control parameter (or Lagrange multiplier) *ϑi* . Fixing the values of {*ϑi*} is equivalent to fix the mean values of {*ki*}. In order to further clarify the role of {*ϑi*} in controlling the topology, let us define in general *xi* = e−*ϑi* . From aprevious equation, modified for directed networks, we have now two kinds of control parameters {*xi, yj* } and so *xiyj* = e−Θi e−Θj . Now, knowing the set {*ϑ*} for all nodes, the ensemble is such that, for each network in Ω, two nodes *i* and *j* are directly connected with a probability given by

$$\mathrm{p}_{i\to j} = \frac{x_i\, y_j}{1 + x_i\, y_j}$$

where *xi*(*yi*) quantifies the ability of node i to receive incoming (form outgoing) connections. Suppose to have incomplete information about the topology of a given network G.

In particular one have the in-degree sequence $\{k_i^{in}\}_{i\in I}$ and the out-degree sequence $\{k_i^{out}\}_{i\in I}$ only for a subset $I \subset G$ of the nodes and conversely a pair of fitness a pair of fitness $\{\chi_j\} \in V$ and $\{\psi_j\} \in V$ where

$$s_i^{in} = \sum_{i\epsilon V} w_{j\to i} \equiv \chi_i \text{ and } s_i^{out} = \sum_{j\epsilon V} w_{i\to j} \equiv \psi_i \text{ for all the nodes.}$$

The CM defines the probability distribution over Ω subject to Lagrange multipliers $\{x_i, y_i\}$ (two for each node), whose values must satisfy the equivalence $< k_i^{in} > = k_i^{in}$ and $< k_i^{out} > = k_i^{out}$, ∀i.

The assumption that the fitnesses χ_i *and* ψ_i are proportional to the in-degree-induced and out-degree-induced Lagrange multipliers $\{x_i\}$ and $\{y_i\}$ through universal (unknown) parameters *α* and *β*:

$x_i \equiv \sqrt{\alpha\chi_i}$ and $y_i \equiv \sqrt{\alpha\psi_i}$ leads to

$$\mathrm{p}_{i\to j} = \frac{\sqrt{\alpha\chi_j}\sqrt{\alpha\psi_i}}{1+\sqrt{\alpha\chi_j}\sqrt{\alpha\psi_i}} = \frac{z\,\chi_j\psi_i}{1+z\,\chi_j\psi_i}$$

where z≡ $\sqrt{\alpha\beta}$. We can now impose the condition that ensures the likelihood

$$\sum_{i\epsilon I}[< k_i^{in} >_\Omega + < k_i^{out} >_\Omega] = \sum_{i\epsilon I}[k_i^{in} + k_i^{out}]$$

where $< k_i^{in} >_\Omega = \sum_{j(\neq i)} p_{j\to i}$ and $< k_i^{out} >_\Omega = \sum_{j(\neq i)} p_{i\to j}$ so now we have an algebraic equation in z to solve numerically

$$\sum_{i\epsilon V}\sum_{j\neq i}\left[\frac{z\chi_i\psi_j}{1+z\chi_i\psi_j} + \frac{z\chi_j\psi_i}{1+z\chi_j\psi_i}\right] = \sum_{i\epsilon V}[k_i^{in} + k_i^{out}]$$

Using z and the fitnesses $\{\chi_i, \psi_i\}$, we generate the ensemble *Ω* by placing a direct link from *i* to *j* with probability $\mathrm{p}_{i\to j}$ with a weight

$$\mathrm{w}_{i\to j} = \frac{\chi_j\psi_i \mathrm{a}_{i\to j}}{\sum_{l,m\epsilon V}\chi_m\psi_l a_{l\to m}} W$$

Where $W = \sum_i \chi_i$, and l and m are referred to the edges present in the bootstrapped network with the relative strength in the real net. Then, we compute the estimate of the variable of interest X(g) on the networks ensemble *Ω* as $< X >_\Omega \pm \sigma_X^\Omega$ averaging the results.

Appendix 3: Vector F[1]:

Fund - hypothetical purchasing of NPLs asset classes expressed in thousand of Euro. It has been calculated dividing the total exposure 13.6bn by the percentual exposure of funds listed in Quaestio - Quaestio (2016). This last data were calculated as usual, summing the exposure listed in Quaestio, and dividing each single exposure by the sum.

Anacap	2583430
Fortress	2583430
Prelios	2311490
Deutsche bank	1541899,8
Eurocastle	1284916,5
Pimco	770949,9
GWM	770949,9
Pve Capital	770949,9
Cerberus	726079,8
Lone Star	648576,9
Credito Fondiario	513966,6
TPG	403830,9
Algebris	285537
Morgan Stanley	142768,5
Poste Vita	142768,5
Baml	142768,5
Beni Stabili	142768,5
GS	141408,8
Ares Managment	135970

Correlation Matrix C6 between origin sectors:

	GOV	ADM	FIN	INS	NFF	OTH
GOV	1	0,6571	0,5612	0,4498	0,5850	0,6423
ADM	0,6571	1	0,4985	0,1742	0,6637	0,5461
FIN	0,5612	0,4985	1	0,6533	0,6475	0,7125
INS	0,4498	0,1742	0,6533	1	0,5920	0,7640
NFF	0,5850	0,6637	0,6475	0,5920	1	0,9215
OTH	0,6423	0,5461	0,7125	0,7640	0,9215	1

Appendix 4: Supplementary figures

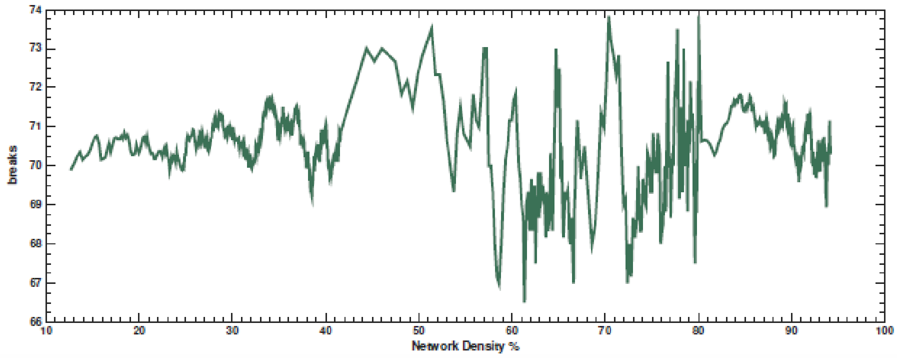


Figure 9. Frequencies of breaks of F 1 for q = 80%

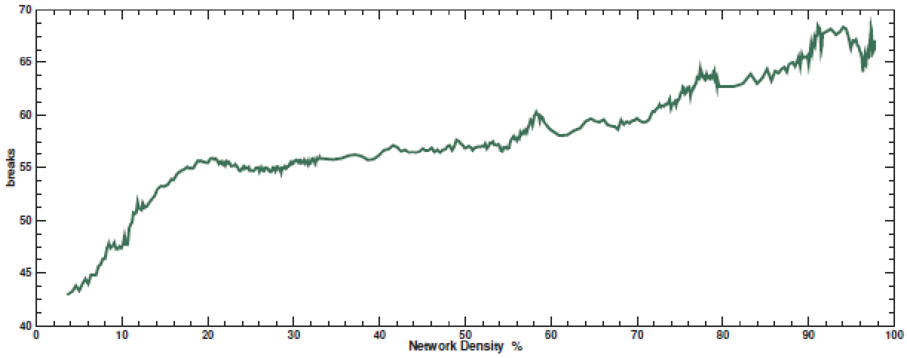


Figure 10. Frequencies of breaks of F 1for q = 50%

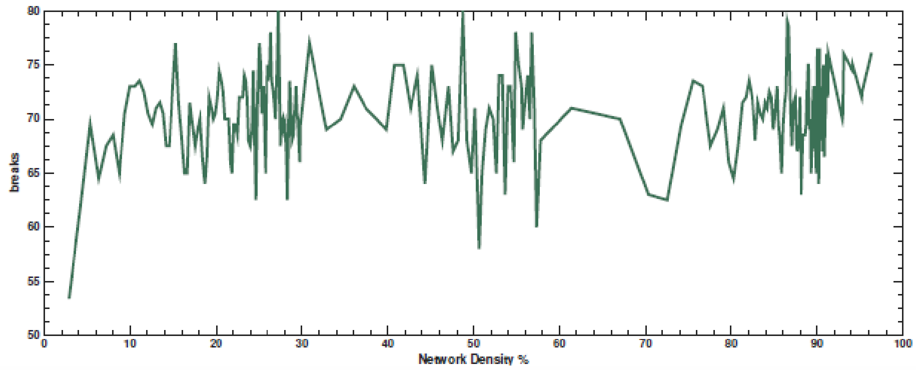


Figure 11. Frequencies of breaks of F 2 for q = 80%

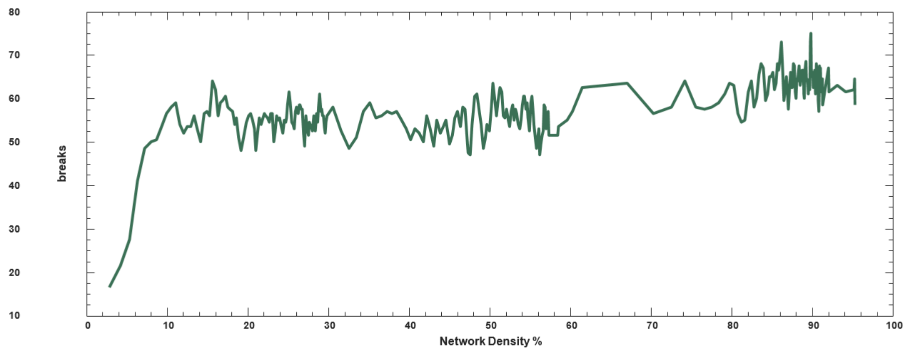


Figure 12. Frequencies of breaks of F 2 for q = 50%

CAUSAL RELATIONSHIP BETWEEN CURRENT ACCOUNT DEFICIT RATE AND INDUSTRIAL PRODUCTION GROWTH IN TURKEY: MIXED-FREQUENCY VAR APPROACH

Rahmi YAMAK
Prof. Dr., Faculty of Economics and Administrative Sciences,
Karadeniz Technical University, Trabzon, Turkey

E-mail: yamak@ktu.edu.tr

Nebiye YAMAK
Prof. Dr., Faculty of Economics and Administrative Sciences,
Karadeniz Technical University, Trabzon, Turkey

E-mail: nyamak@ktu.edu.tr

Serkan SAMUT
Research Assistant, Faculty of Economics and Administrative Sciences,
Karadeniz Technical University, Trabzon, Turkey

E-mail: serkan.samut@ktu.edu.tr

Abstract
In this study, the causal relationship between current account deficit rate and industrial production growth in Turkey is investigated by applying the mixed-frequency VAR approach developed by Ghysels et al. (2016). The data span from January 2005 to September 2018 for monthly industrial production index and from the first quarter of 2005 to the third quarter of 2018 for the quarterly current account deficit rate. Granger causality tests suggest that there is only one-way causal relationship between in the variables. It is running from industrial production to current account. Impulse response functions confirm the same causal relationship. Based on causality tests and impulse response functions, it can be concluded that the main reason for the recent improvement in the current account deficit is the contraction in industrial production in Turkey.

Keywords: *Current Account Deficit; Industrial Production; M-F VAR; Granger Causality; Turkish Economy*

1. Introduction

Over the last two decades, one of the two most discussed macroeconomic indicators on the Turkish economy is the current account deficit rate and the other is the industrial production growth. The current account deficit to GDP ratio in Turkey had an average of 5.5% between the years 2005-2018. It was in a continuous recovery period by fluctuating after reaching its highest level in the first quarter of 2011. Even in the third quarter of 2018, the current account yielded a surplus of 1%. Despite the positive developments in the current account deficit, industrial production experienced serious contradiction in the same period. Thus, in the period of 2005-2018, industrial production grew only by 1% on a monthly average and decreased by 9.8% in December 2018 having the sharpest decline in the last 10 years. The recent developments in both the current account deficit and the industrial production in Turkey have led to a re-evaluation of the relations between these two important indicators. Although the recent improvement in the current account deficit was seen as an economic success, the real sector of the economy argues that this is not a success but rather a result of the contraction in industrial production. Over twenty years, industrial production in Turkey is dependent on imports of intermediate and capital goods. Since the economic policy has been based on the model called as first import then export, it has been argued that the main reason for the improvement in the current account deficit is the contraction in industrial production. Another view argues that the decline in industrial production is not due to the foreign trade model, on the contrary, due to the contraction in domestic demand as well as stock surplus.

The related empirical literature is rich in terms of the study investigating the causal relationships between the current account deficit and economic growth. There are some studies that determine the bi-directional causal relationship as well as studies that determine one-way causal relationship between the two variables. Sadaf & Amin (2018), Akbaş *et al.* (2014), Erdoğan & Acet (2016), Arslan *et al.* (2017) can be given as an example of studies which find a bi-directional causal relationship. Malik *et al.* (2010), Erataş (2014) and Özer *et al.* (2018) are some of the studies that determine one-way causal relationship from current account deficit rate to economic growth. Uçak (2017), Yurdakul & Ucar (2015), Duman (2017), Yılmaz & Akıncı (2011) are among the studies detecting one-way causality from economic growth to current account deficit.

Although the applied method and the analyzed country in the studies differed, they all have a common feature. It is that they used either quarterly or annual data in their econometric analyses. In the current literature, the main reason for not using the high frequency series as monthly is the lack of high frequency of the GDP series, which is required in producing the current account balance rate series. Although the current account balance is available on a monthly basis, since the GDP is not present, quarterly data have been used as the highest frequency. As known, in the traditional multi time series analysis it is a necessity for the variables to be on the same frequency. In a two-variable time series analysis, it is not possible to use one of the variables at low frequency and the other at high frequency. However, until the 2000s, the most effective method used to eliminate the frequency difference in the series was to convert the high frequency series to the frequency of the lowest frequency variable in the analysis. The conversion was carried out by naturally aggregating the high frequency series for the desired frequency. However, such an approach may cause the findings to be statistically inefficient and biased (Andreou *et al.,* 2010). In addition, high frequency series contains more information than low frequency. As a result of temporal aggregation, the potential information in the high frequency series either disappears or has a different distribution (Marcellino, 1999; Götz *et al.,* 2015). Therefore, reducing the frequency of any time series with this approach means allowing the loss of potential information in the series. Granger (1988), Pesaran *et al.* (1989), Granger & Siklos (1995) demonstrated that the results of the analysis performed with temporally aggregated variables may be different from the findings obtained by disaggregated variables. For this reason, there are some doubts about the statistical reliability and accuracy of the analysis results with temporal aggregated variables.

The related frequency literature has drawn attention to the negative effects of the temporal aggregation approach since the beginning of the 2000s and hence focused on solution-oriented studies. One of the pioneering works in this area is Ghysels *et al.* (2004). In their study, Ghysels *et al.* (2004) developed a method called Mixed-Data Sampling (MIDAS), which allows a combination of different frequency series. This approach is a single-equation model approach in which low-frequency dependent variable and high-frequency independent variables or variables are used together. Later on, Ghysels (2016) and Ghysels *et al.* (2016) developed the Mixed-Frequency VAR (MF-VAR) method for VAR analysis, where all variables areassumed to endogenous. Therefore, the MF-VAR approach allows for Granger causality testing between different frequency series.

The purpose of the present study is to detect the possible causal relationship between current account deficit rate and industrial production growth in the light of the above discussions for the case of Turkey. In this study, the causal relationships between the quarterly current account deficit rate and the monthly industrial production growth were analyzed by using the MF-VAR approach for the period 2015-2017.

2. Data Set and Method

The data span from January 2005 to September 2018 for monthly industrial production index (IP) and from the first quarter of 2005 to the third quarter of 2018 for the quarterly current account deficit rate (CADR). All data were obtained from Central Bank of the Republic of Turkey. In this study, industrial production growth rates (DLIP) were calculated from logarithmic industrial production data (LIP). As mentioned earlier, the possible causality between CADR and DLIP is revealed by Granger causality under MF-VAR method which allows analysis on the original frequencies of the variables adapted for causality. The MF-VAR model is as shown in equation (2.1).

$$X(\tau_L) = (DLIP(\tau_L, 1)' \dots\dots .. DLIP(\tau_L, m)', CADR(\tau_l)')' \qquad (2.1)$$

In equation (2.1) above, $X(\tau_L)$ represents both high frequency DLIP variable and low frequency CADR variable. τ_L is low frequency time period while m is the number of time periods with high frequency corresponding to a time unit of the low frequency variable. In the MF-VAR model with $p \geq 1$, it is assumed that $X(\tau_L)$ follows the MF-VAR (p) process.

$$X(\tau_L) = \sum_{k=1}^{p} \beta_k X(\tau_L - k) + \varepsilon(\tau_L) \qquad (2.2)$$

In the above MF-VAR (p) model; $k = 1, \dots\dots. p$ is the coefficient matrix and $\varepsilon(\tau_L)$ is the error vector. Here, there is an assumption that error vector with $k \times 1$ follows a stationary process with $\varepsilon(\tau_L)$ =$[\varepsilon_1(\tau_L), \dots\dots.., \varepsilon_k(\tau_L)]'$

The least squares method is used to estimate the MF-VAR model. However, in calculating the variance-covariance matrix for the parameters, Newey & West (1987)'s HAC variance estimator and Newey & West (1994)'s automatic lag selection are used. Then, it is decided whether there is a causal relationship between variables by calculating the Wald statistics.

3. Findings

In this study, firstly the stationarity of CADR and LIP was examined by using the Augmented Dickey-Fuller (ADF) unit root test approach. In Table 1, it can be seen that CADR is stationary on its level while LIP is stationary on its first differences. These results mean that both current account deficit rate and industrial production growth are stationary on their levels.

Table 1. ADF Unit-Root Test Results

Variable	Constant	Constant and Trend
CADR	-2.946**	-2.899
LIP	-0.551	-2.073
DLIP	-6.587***	-6.522***

Note: ***, ** and * indicate 1%, 5% and 10% significance, respectively. CADR; The current account deficit rate which is calculated by dividing the GDP by the current account deficit, LIP; Logarithmic Industrial Production Index, D shows that the first difference of the series.

In the empirical analyze, MATLAB codes written by Motegi (2014) were used to perform Granger causality test based on the MF-VAR model between CADR and DLIP. In the MF-VAR model, monthly data of the high-frequency industrial production growth rate is divided into three sections, which correspond to the frequency of the quarterly current account deficit rate series. In the separation process, industrial production growth rates corresponding to the first months of the quarterly periods, the second months in the second part and the last months in the third part are included. Therefore, in the MF-VAR model, Granger causality test is applied on four different variables instead of two variables. Since there is no fixed term in the MF-VAR model developed by Ghysels *et al.* (2016), deviations of the variables from their averages are used in the analyzes.

Granger causality test was applied first between the monthly DLIP and quarterly CADR. Then the same test was also implemented between quarterly aggregated DLIP and quarterly CADR. The optimal lag length for both causality tests was determined to be one by using Final Prediction Error criteria. Table 2 reports the probabilities for rejecting the null hypotheses that there are no causal relationships between the two variables. The null hypothesis that implies no causal relationship from CADR to DLIP cannot be rejected in both mixed frequency and low frequency analyses. The probabilities for rejecting the null hypothesis are 0.173 and 0.589, respectively. On the other hand, according to probabilities given in the same table, the null hypothesis that indicates no causal relationship from DLIP to CADR is rejected in both mixed frequency and low frequency analyses. The probability for rejecting the null hypothesis is 0.004 in mixed frequency analysis while it is 0.019 in low frequency approach. According to both analyses, there is a causal relationship between both variables for the case of Turkey. It is running from industrial production growth rate to current account deficit rate. However, this conclusion is stronger in mixed-frequency (MF-VAR) model than in low frequency model.

Table 2. Probability Values of Granger Causality Test

H_0 Hypothesis	Mixed Frequency	Low Frequency
CADR "$\nRightarrow$" **DLIP**	0.173 (1)	0.589 (1)
DLIP "$\nRightarrow$" **CADR**	0.004 (1)	0.019 (1)

Note: The values in parentheses are the optimal lag length for MF-VAR and traditional VAR Models.

The diagnostic test statistics of MF-VAR and traditional VAR (aggregated industrial production growth rate) are given in Table 3. From the table, it can be observed that that there is no autocorrelation problem in both models. In addition, White heteroskedasticity test indicates that there is no heteroskedasticity problem in the MF-VAR model. But, the null hypothesis that implies the non-existence of the heteroskedasticity problem can be rejected at the significance level of 0.05 for the traditional VAR model. In other words, there exists a strong heteroskedasticity problem in the traditional VAR model where the monthly industrial production growth rate is converted into three months. AR roots of MF-VAR and traditional VAR models are presented in Graph 1. As can be seen from the graph, there is no unit root in both models.

Table 3. Diagnostic Test Results

	Mixed Frequency	**Low Frequency**
LM Autocorrelation Test	14.573 (0.556)	5.993 (0.2)
White Heteroskedasticity Test	139.318 (0.5)	29.558 (0.014)

The values in brackets are probability values of the calculated statistics.

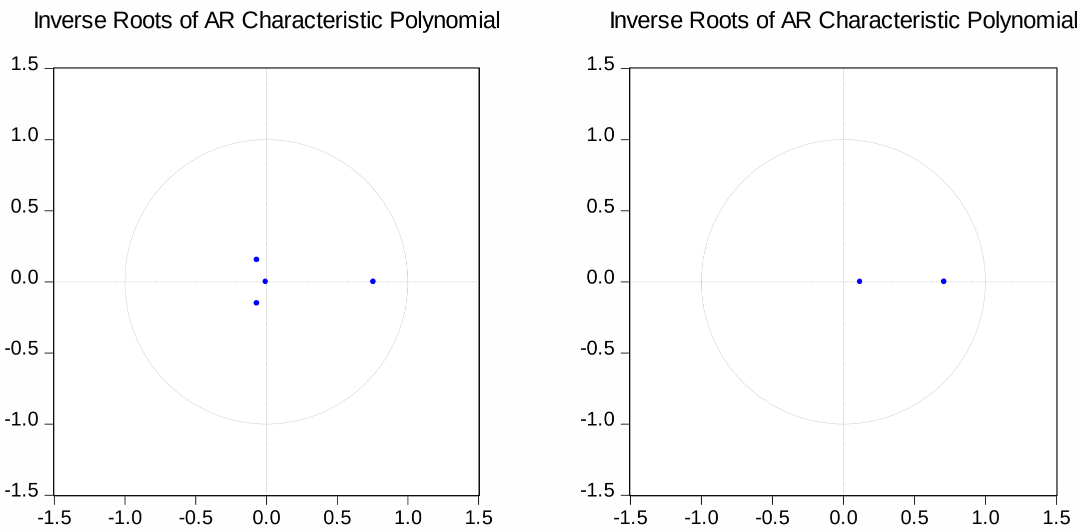


Figure 1. AR Roots for MF-VAR and Traditional VAR Models

Figure 2 shows the impulse-response functions of the MF-VAR model. In the figure, DLIP1, DLIP2 and DLIP3 are quarterly industrial production growth equivalent to the 1st, 2nd and 3rd months respectively. The first three graphs at the last column of Figure 2 show the response of DLIP to a one standard deviations shock to CADR for the 1st, 2nd, and 3rd months, respectively. In the first quarter, the response of DLIP to a shock in CADR is statistically significant and slightly negative for 1st and 3rd months, and slightly positive for 2nd month. After the first quarter, the response is no longer significant and after that, the impulse response function stabilizes around zero. The degree of the response of the industrial production growth rate to the current account deficit seems to be consistent with the above Granger causality test results. As can be recalled, the current deficit rate does not Granger-cause industrial production growth rate. Here too, industrial production growth rate does not give a significant response to any shock in current account deficit rate.

On the other hand, the first three graphs at the last row of Figure 2 demonstrate the response of current account deficit rate to a one standard deviations shock to industrial production growth for the 1st, 2nd, and 3rd months, respectively. The response of the current account deficit to the shock of industrial production growth shows a similar pattern for three months in the quarter. As can be seen from the figure, the response of the current deficit rate is similar to the V shape. As a result of the shock of industrial production growth, the current account deficit rate is first increasing, then decreasing and approaching the zero balance current account. The response of current account deficit rate to a standard deviation shock in the industrial production growth reaches its maximum value in the first quarter. The response of current account deficit rate disappears by approaching the zero value towards the fifth quarter. The V shape on the response of current account deficit rate to industrial production shock means that industrial production growth rate Granger-causes current account deficit rate. This conclusion is also consistent with the probability value of Granger causality test.

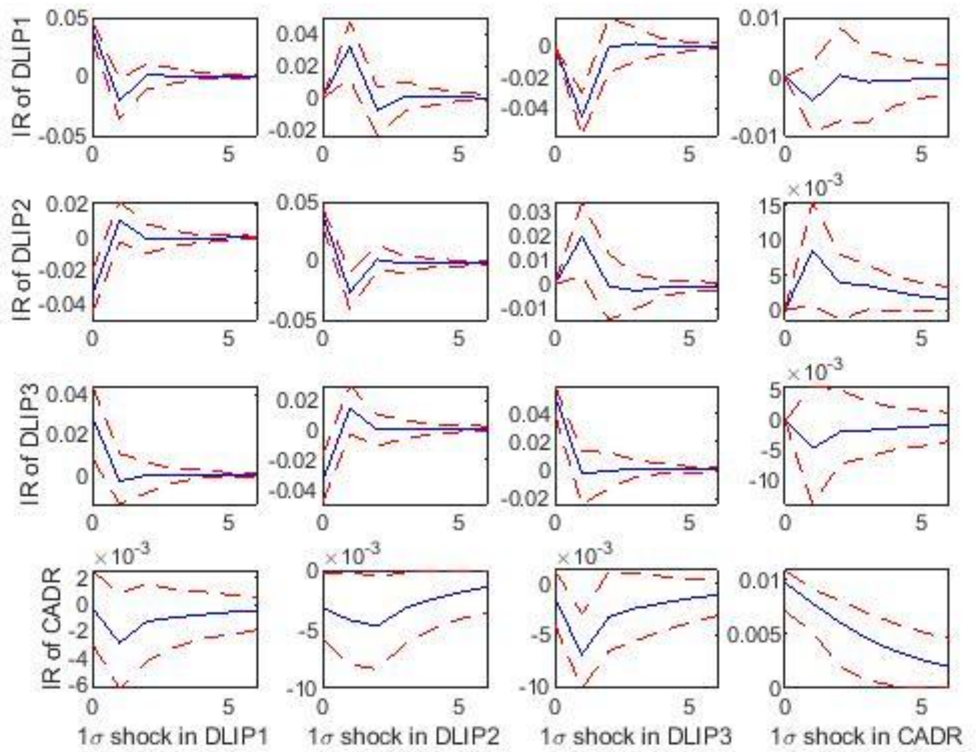


Figure 2. MF-VAR Impulse Response Function

As previously mentioned, in this present study, causality relationship between the two variables was investigated by using traditional VAR as well as MF-VAR. In Figure 3, impulse-response functions of traditional VAR model are presented. The graph at the top right of the relevant figure shows the response of aggregated industrial production growth to a one standard deviations shock to current account deficit rate. The response of industrial production growth to a shock in current account deficit rate is statistically insignificant and the impulse-response function stabilizes around zero by indicating that current account deficit rate dos not Granger-cause industrial production growth rate. The graph in the lower left part of

Figure 3 demonstrates the response of current account deficit rate to a one standard deviations shock to aggregated industrial production growth. The initial response of current account deficit rate to a shock in industrial production growth is statistically significant and negative. This negative shock reaches the maximum value in the first quarter. After that, it is approaching zero by slowly decreasing.

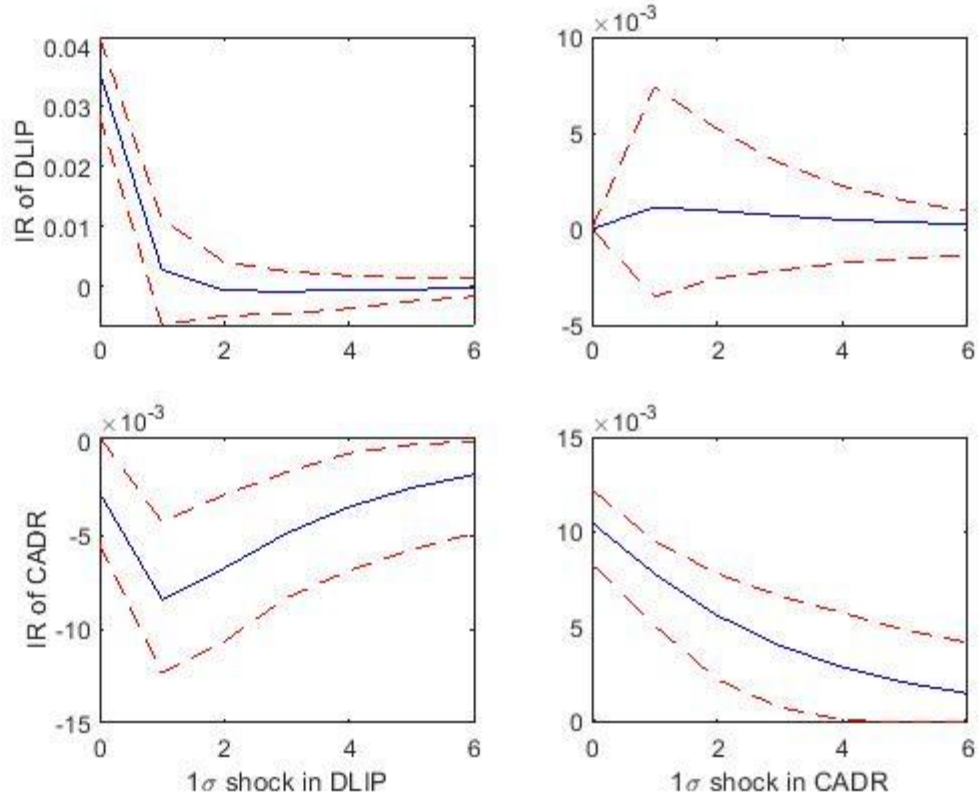


Figure 3. Traditional VAR Impulse Response Function

4. Conclusion

In this study, the possible causal relationship between current account deficit rate and industrial production growth rate were investigated by using Mixed-frequency VAR approach developed by Ghysels *et al.* (2016). The biggest advantage of the MF-VAR model is that it allows Granger causality test between the series with different frequency. The aim of the study is to determine whether there is any causal relationship between monthly DLIP and quarterly CADR for the case of Turkey.

Initially, both conventional and MF-VAR models were estimated separately. Then, Granger causality tests were performed based on both models. The findings of traditional VAR indicate that there is only one-way causality from quarterly industrial production growth rate to quarterly current account rate. The same result was also obtained from the MF-VAR model. According to MF-VAR, monthly industrial production growth rate Granger-causes quarterly current account deficit rate. There is no difference between the two models in terms of determining the direction of the causal relationship. However, the one-way causal relationship to the current account deficit from industrial production growth rate was found to be statistically stronger in the MF-VAR model, where high-frequency series was employed, than in traditional VAR model. Impulse response functions of MF-VAR validated one-way causality from DLIP to CADR. The response of CADR to DLIP seems to be V shape. In addition, the hypothesis that there is no causal relationship between the current deficit ratio and the industrial production growth rate in both models could not be rejected. However, in the MF-VAR model where the high frequency series is used, the probability of rejecting no Granger causality from current account deficit rate to industrial production growth rate is higher than in the conventional model.

From the empirical analysis of the study, it can be seen that the recent improvements in the Turkish current account deficit are not the result of a policy performance. On the contrary, these improvements seem to be a result of the contraction in industrial production. Whenever industrial production growth rate starts to increase, the current account deficit rate is likely to increase.

References

1. Akbaş, Y. E., Lebe, F. & Uluyol, O. **Analyzing The Relationship Among The GDP-Current Account Deficit and Short Term Capital Flows: The Case of Emerging Markets**, Journal of Management & Economics, Vol. 21, Issue 2, 2014, pp. 293-303.
2. Andreou, E., Ghysels, E. & Kourtellos, A. **Regression Models with Mixed Sampling Frequencies, Journal of Econometrics,** Vol. 158, Issue 2, 2010, pp. 246-261.
3. Arslan, İ., Uğur, A.A. & Dineri, A. **Relationship between Current Account Deficit and Economic Growth in OECD Countries: 1980-2014**, Journal of Economics Business and Political Researches, Vol. 2, Issue 3, 2017, pp. 56-68.
4. Duman, Y. K. **The Relationship Between Current Account Balance And Economic Growth in Turkey** (in Turkish), Research of Financial Economic and Social Studies (RFES), Vol. 2, Issue 4, 2017, pp. 231-244.
5. Erataş, F., **Panel Causality Analysis of Relationship Between Current Account Deficit And Economic Growth: Example of the G7 Countries**, International Conference in Economics Prague, Czech Republic, September 03-05, 2014
6. Erdoğan, S. & Acet, H. **The Relationship Between Current Account Balance and Economic Growth: Turkey Case (2003-2015)** (in Turkish), The Journal of Academic Social Science Studies, Vol. 51, 2016, pp. 539-548.
7. Ghysels, E. **Macroeconomics and the Reality of Mixed Frequency Data**, Journal of Econometrics, Vol. 193, 2016, pp. 293-314.
8. Ghysels, E., Hill, J.B. & Motegi, K. **Testing for Granger Causality with Mixed Frequency Data**, Journal of Econometrics, Vol. 192, Issue 1, 2016, pp. 207-230.
9. Ghysels, E., Santa-Clara, P. & Valkanov, R. **The MIDAS Touch: Mixed Data Sampling Regression Models**, https://cloudfront.escholarship.org/dist/prd/ content/qt9mf223rs/ qt9mf223rs. Pdf, 2004.
10. Götz, T. B., Hecq, A. & Smeekes, S. **Testing for Granger Causality in Large Mixed-Frequency VARs**, Discussion Paper, Deutsche Bundesbank. No. 45, 2015.

11. Granger, C. W. J. **Aggregation of Time Series Variables: A Survey**, Institute for Empirical Macroeconomics Federal Reserve Bank of Minneapolis Discussion Paper, Vol. 1, 1988, pp. 1-25.
12. Granger, C. W. J. & Siklos, P.L. **Systematic Sampling, Temporal Aggregation, Seasonal Adjustment, and Cointegration Theory and Evidence**, Journal of Econometrics, Vol. 66, Issues 1-2, 1995, pp. 357-369.
13. Malik, S., Chaudhry, I.S., Sheikh, M.R. & Farooqi, F.S. **Tourism, Economic Growth and Current Account Deficit in Pakistan: Evidence from Co-integration and Causal Analysis**, European Journal of Economics, Finance and Administrative Sciences, Vol. 22, 2010, pp. 21-31.
14. Marcellino, M. **Some Consequences of Temporal Aggregation in Empirical Analysis**, Journal of Business & Economic Statistics, 17 (1), 1999, 129-136.
15. Motegi, K. **MFVAR Toolbox**, (15.06.2018). http://www.kobe-u.ac.jp/~motegi/Matlab_Codes.html, 2014.
16. Newey, W. K. & West, K. D. **A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix**, Econometrica, Vol. 55, 1987, pp. 703–708.
17. Newey, W. K. & West, K.D. **Automatic Lag Selection in Covariance Matrix Estimation**, The Review of Economic Studies, Vol. 61, Issue 4, 1994, pp. 631-653.
18. Özer, M., Žugić, J. & Tomaš-Miskin. S. **The Relationship Between Current Account Deficits and Growth in Montenegro: ARDL Bounds Testing Approach**, Journal of Central Banking Theory and Practice, Vol. 7, Issue 3, 2018, pp. 5-24.
19. Pesaran, M. H., Pierse, R.G. & Kumar, M.S. **Econometric Analysis of Aggregation in the Context of Linear Prediction Models**, Econometrica: Journal of the Econometric Society, Vol. 57, 1989, pp. 861-888.
20. Sadaf, B.H. & Amin, S.B. **Causal Relationship between Current Account Deficit and Foreign Direct Investment: An Empirical Analysis of Bangladesh**, World Review of Business Research, Vol. 8, Issue 3, 2018, pp. 52-60.
21. Uçak, S. **The relationship between Current Account Balance and Economic Growth: The Analyses of Turkey** (in Turkish), Journal of Social Sciences, Vol. 15, Issue 2, 2017, pp. 107-140.
22. Yılmaz, Ö. & Akıncı, M. **The Relationship Between Economic Growth and Current Account Balance: The Case of Turkey** (in Turkish), Journal of Graduate School of Social Sciences, Vol. 15, Issue 2, 2011, pp. 363-377.
23. Yurdakul, F. & Ucar, U. **The Relationship Between Current Deficit and Economic Growth: An Empirical Study on Turkey,** Procedia Economics and Finance, Vol. 26, 2015, pp. 101-108.

A QUASI GARIMA DISTRIBUTION

Rama SHANKER
Department of Statistics,
Eritrea Institute of Technology, Asmara, Eritrea

E-mail: shankerrama2009@gmail.com

Kamlesh Kumar SHUKLA
Department of Statistics,
Eritrea Institute of Technology, Asmara, Eritrea

E-mail: kkshukla22@gmail.com

Simon SIUM
Department of Statistics,
Eritrea Institute of Technology, Asmara, Eritrea

Daniel WELDAY
Department of Statistics,
Eritrea Institute of Technology, Asmara, Eritrea

Ravi SHANKER
Department of Mathematics,
G.L.A. College, N.P University, Daltonganj, Jharkhand, India

E-mail: ravi.shanker74@gmail.com

Abstract

In the present paper, a two-parameter quasi Garima distribution (QGD) which includes one parameter exponential distribution and Garima distribution introduced by Shanker (2016 c) as special cases, has been proposed. Its statistical and mathematical properties including moments and moments based measures, hazard rate function, mean residual life function, stochastic ordering, mean deviations, Bonferroni and Lorenz curves, order statistics, Renyi entropy measure and stress-strength reliability have also been discussed. The method of moments and the method of maximum likelihood estimation have been discussed for estimating the parameters of QGD. Finally, the goodness of fit of the QGD has been discussed with a real lifetime dataset and the fit is quite satisfactory over one parameter and two-parameter lifetime distributions.

Keywords: *Garima distribution; Moments; Reliability Properties; Stochastic ordering; Mean deviations; Stress-strength reliability; Estimation of parameters; goodness of fit*

1. Introduction

Shanker (2016 c) has introduced a one parameter lifetime distribution named Garima distribution for modeling lifetime data from behavioral science having probability density function (pdf) and cumulative distribution function (cdf)

$$f_1(x;\theta)=\frac{\theta}{\theta+2}(1+\theta+\theta x)e^{-\theta x} \quad ;x>0,\ \theta>0. \tag{1.1}$$

$$F_1(x,\theta)=1-\left[1+\frac{\theta x}{\theta+2}\right]e^{-\theta x} \ ;x>0,\theta>0 \tag{1.2}$$

Shanker (2016 c) has shown that it gives better fit than one parameter exponential distribution, Lindley distribution introduced by Lindley (1958) and Shanker, Akash, Aradhana and Sujatha distributions introduced by Shanker (2015 a, 2015 b, 2016 a, 2016 b). This distribution is a convex combination of exponential (θ) and gamma $(2,\theta)$ distributions with their mixing proportion $\frac{\theta+1}{\theta+2}$.

The first four moments about origin of Garima distribution obtained by Shanker (2016 c) are given as

$$\mu_1' = \frac{\theta+3}{\theta(\theta+2)}, \quad \mu_2' = \frac{2(\theta+4)}{\theta^2(\theta+2)}, \quad \mu_3' = \frac{6(\theta+5)}{\theta^3(\theta+2)}, \quad \mu_4' = \frac{24(\theta+6)}{\theta^4(\theta+2)}$$

The central moments of Garima distribution obtained by Shanker (2016 c) are

$$\mu_2 = \frac{\theta^2+6\theta+7}{\theta^2(\theta+2)^2}$$

$$\mu_3 = \frac{2(\theta^3+9\theta^2+21\theta+15)}{\theta^3(\theta+2)^3}$$

$$\mu_4 = \frac{3(3\theta^4+36\theta^3+134\theta^2+204\theta+111)}{\theta^4(\theta+2)^4}$$

Shanker (2016 c) studied its important properties including coefficient of variation, skewness, kurtosis, Index of dispersion, hazard rate function, mean residual life function, stochastic ordering, mean deviations, order statistics, Bonferroni and Lorenz curves, Renyi entropy measure, and stress-strength reliability. Shanker (2016 c) has also discussed the estimation of parameter using both the method of moments and the method of maximum likelihood estimation and the application of the distribution to model behavioral science data. The discrete Poisson – Garima distribution, a Poisson mixture of Garima distribution has also been studied by Shanker (2017).

In this paper, a two - parameter quasi Garima distribution (QGD), of which one parameter exponential distribution and Garima distribution introduced by Shanker (2016 c) are particular cases, has been proposed. Its raw moments and central moments have been obtained and coefficients of variation, skewness, kurtosis and index of dispersion have been discussed. Some of its important mathematical and statistical properties including hazard rate function, mean residual life function, stochastic ordering, mean deviations, Bonferroni and Lorenz curves, order statistics, Renyi entropy measure and stress-strength reliability have also been discussed. The estimation of the parameters has been discussed using both the method of moments and the maximum likelihood estimation. The goodness of fit of QGD has been illustrated with a real lifetime dataset and the fit has been compared with well known one parameter and two-parameter lifetime distributions.

2. A quasi Garima distribution

A two - parameter quasi Garima distribution (QGD) having parameters θ and α is defined by its pdf

$$f_2(x;\theta,\alpha)=\frac{\theta^2}{\theta^2+\theta+\alpha}(1+\theta+\alpha x)e^{-\theta x}\;;x>0,\,\theta>0,\alpha>0. \tag{2.1}$$

It can be easily verified that (2.1) reduces to the exponential distribution and Garima distribution at $\alpha=0$ and $\alpha=\theta$ respectively. It can be easily verified that QGD is a convex combination of exponential (θ) and gamma $(2,\theta)$ distributions. We have

$$f_2(x;\theta,\alpha)=p\,g_1(x;\theta)+(1-p)\,g_2(x;2,\theta) \tag{2.2}$$

where

$$p=\frac{(\theta+1)\theta}{\theta^2+\theta+\alpha},$$

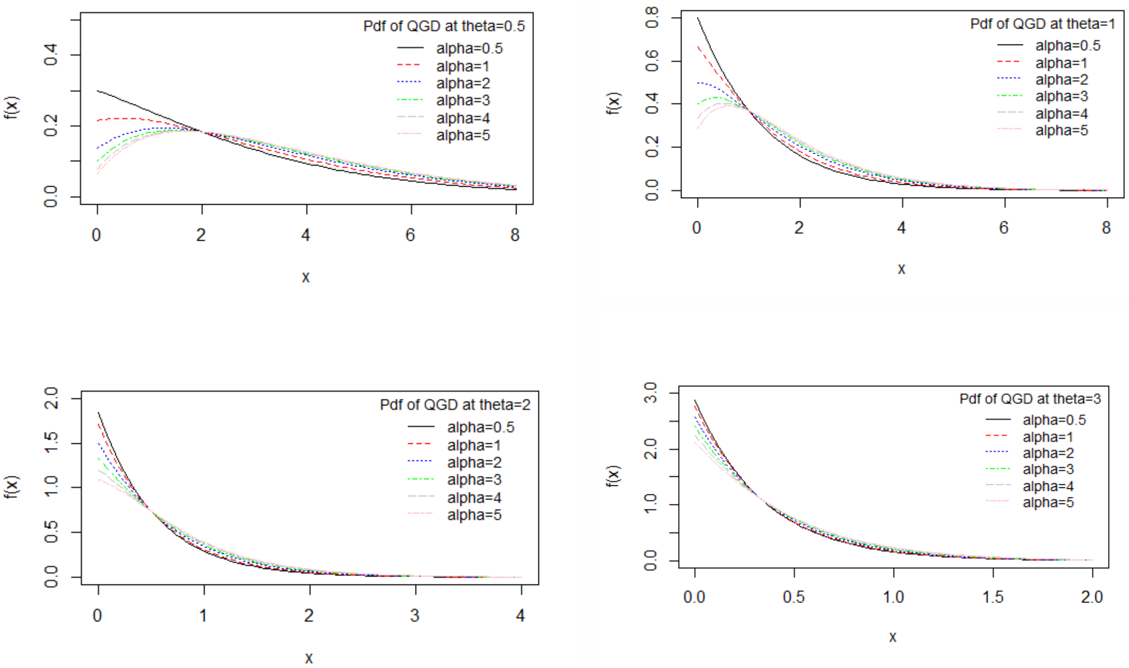
$$g_1(x;\theta)=\theta e^{-\theta x};\;\; x>0,\,\theta>0$$

$$g_2(x;\theta)=\frac{\theta^2}{\Gamma(2)}e^{-\theta x}x^{2-1};\;\; x>0,\,\theta>0.$$

The corresponding cdf of QGD (2.1) can be obtained as

$$F_2(x;\theta,\alpha)=1-\left[1+\frac{\alpha\theta x}{\theta^2+\theta+\alpha}\right]e^{-\theta x};\;\; x>0,\theta>0,\alpha>0 \tag{2.3}$$

The nature and behavior of the pdf and the cdf of QGD for varying values of the parameters θ and α have been explained graphically and presented in figures 1 and 2, respectively. From fig. 1, it is obvious that when θ is fixed and α is changing, there is a slight difference in the shapes of the pdf of QGD. Further, when α is fixed and θ is changing, there is a remarkable difference in the shapes of the pdf of QGD. This means that the parameter θ is playing a dominant role in the shape of the pdf of QGD. The same fact can be observed from the shapes of the cdf of QGD in fig. 2.



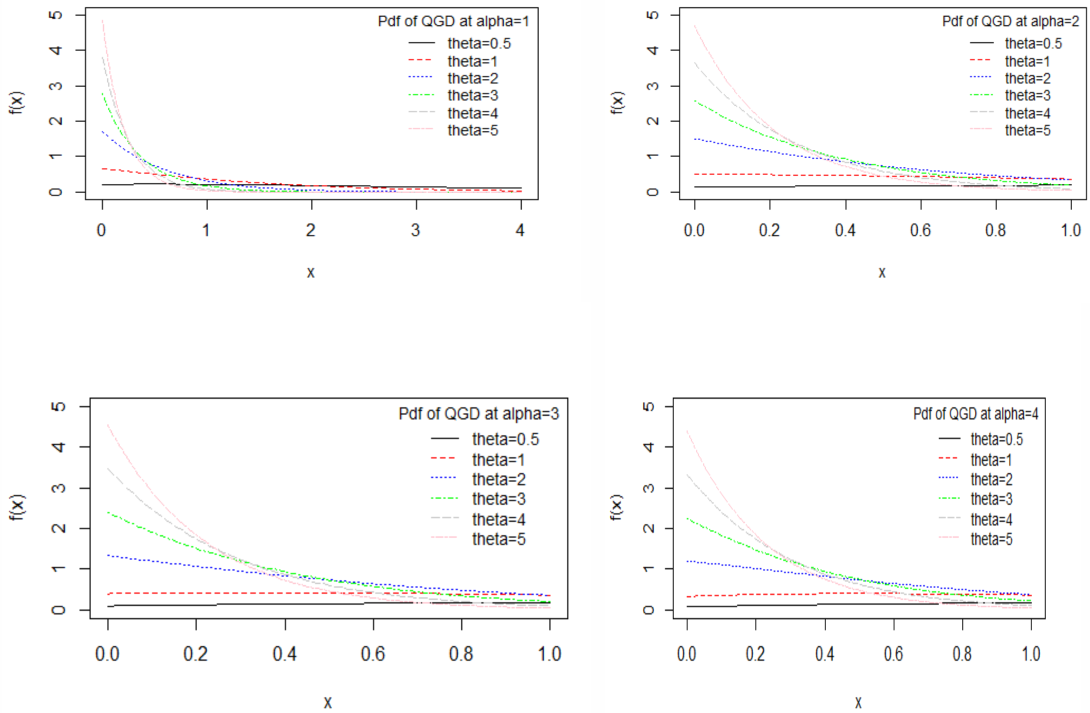
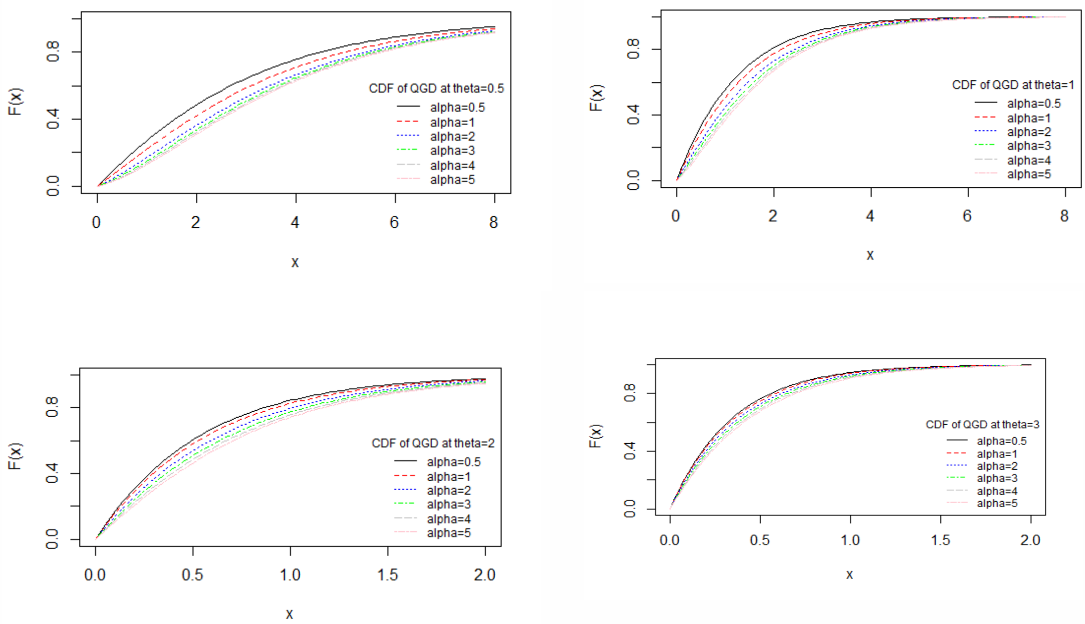


Figure 1. Graphs of the pdf of QGD for varying values of parameters θ and α



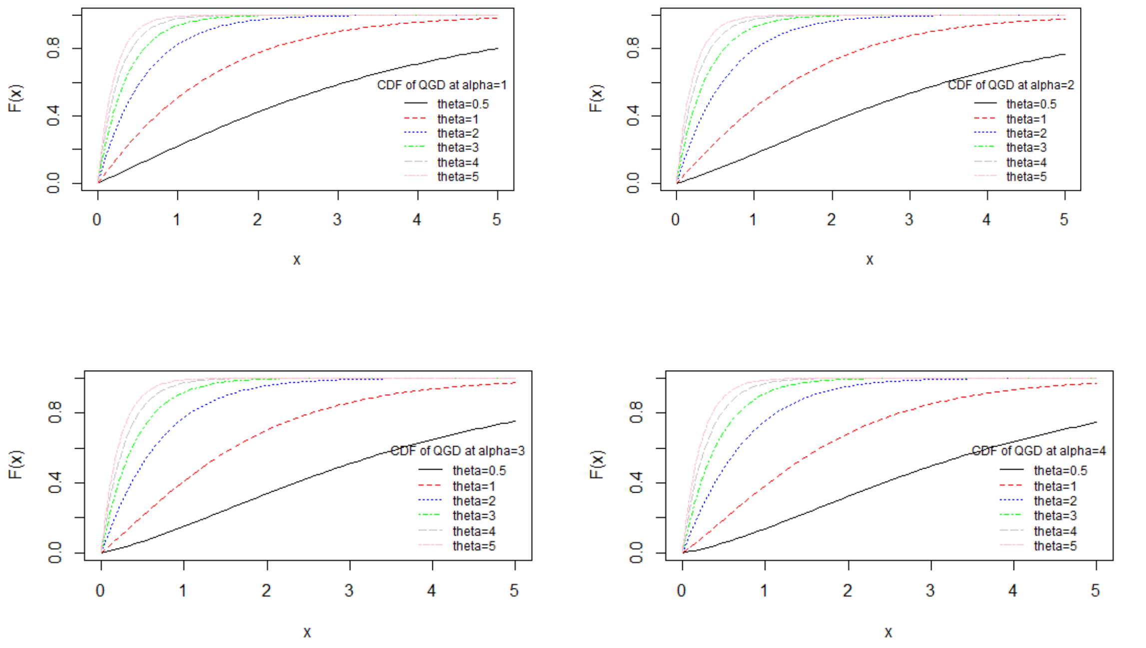


Figure 2. Graphs of the cdf of QGD for varying values of parameters θ and α

3. Statistical constants

Using the convex combination representation (2.2), the r th moment about origin of QGD (2.1) can be obtained as

$$\mu_r' = \frac{r!\left\{\theta^2+\theta+(r+1)\alpha\right\}}{\theta^r\left(\theta^2+\theta+\alpha\right)} \quad ; r=1,2,3,... \tag{3.1}$$

Thus, the first four moments about origin of QGD are given by

$$\mu_1' = \frac{\theta^2+\theta+2\alpha}{\theta\left(\theta^2+\theta+\alpha\right)}, \qquad \mu_2' = \frac{2\left(\theta^2+\theta+3\alpha\right)}{\theta^2\left(\theta^2+\theta+\alpha\right)}$$

$$\mu_3' = \frac{6\left(\theta^2+\theta+4\alpha\right)}{\theta^3\left(\theta^2+\theta+\alpha\right)}, \qquad \mu_4' = \frac{24\left(\theta^2+\theta+5\alpha\right)}{\theta^4\left(\theta^2+\theta+\alpha\right)}$$

Using relationship between central moments and moments about origin, the central moments of QGD are thus obtained as

$$\mu_2 = \frac{\theta^4+2\theta^3+(4\alpha+1)\theta^2+4\theta\alpha+2\alpha^2}{\theta^2\left(\theta^2+\theta+\alpha\right)^2}$$

$$\mu_3 = \frac{2\left\{\theta^6+3\theta^5+3(2\alpha+1)\theta^4+(12\alpha+1)\theta^3+6\alpha(\alpha+1)\theta^2+6\theta\alpha^2+2\alpha^3\right\}}{\theta^3\left(\theta^2+\theta+\alpha\right)^3}$$

$$\mu_4 = \frac{3\begin{Bmatrix} 3\theta^8 + 12\theta^7 + 6(4\alpha+3)\theta^6 + 12(6\alpha+1)\theta^5 + (44\alpha^2 + 72\alpha + 3)\theta^4 \\ +8(11\alpha+3)\theta^3\alpha + 4(8\alpha+11)\theta^2\alpha^2 + 32\theta\alpha^3 + 8\alpha^4 \end{Bmatrix}}{\theta^4\left(\theta^2+\theta+\alpha\right)^4}.$$

The coefficient of variation $(C.V)$, coefficient of skewness $\left(\sqrt{\beta_1}\right)$, coefficient of kurtosis (β_2) and index of dispersion (γ) of QGD are obtained as

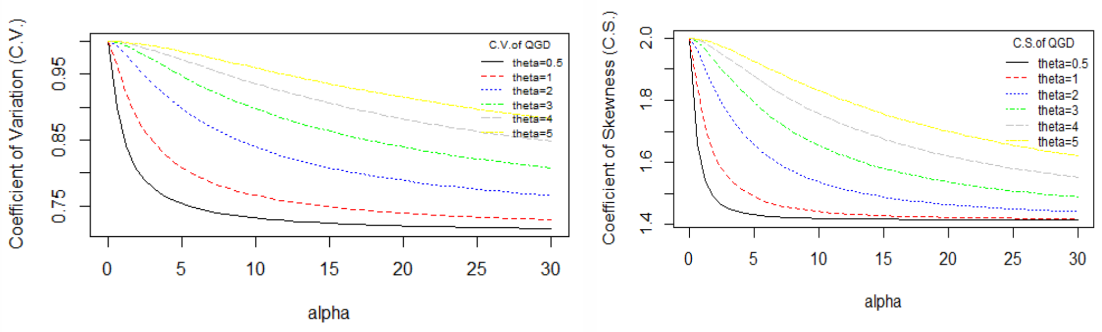
$$C.V = \frac{\sigma}{\mu_1'} = \frac{\sqrt{\theta^4 + 2\theta^3 + (4\alpha+1)\theta^2 + 4\theta\alpha + 2\alpha^2}}{\theta^2+\theta+2\alpha}$$

$$\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{2\left\{\theta^6 + 3\theta^5 + 3(2\alpha+1)\theta^4 + (12\alpha+1)\theta^3 + 6\alpha(\alpha+1)\theta^2 + 6\theta\alpha^2 + 2\alpha^3\right\}}{\left\{\theta^4 + 2\theta^3 + (4\alpha+1)\theta^2 + 4\theta\alpha + 2\alpha^2\right\}^{3/2}}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\begin{Bmatrix} 3\theta^8 + 12\theta^7 + 6(4\alpha+3)\theta^6 + 12(6\alpha+1)\theta^5 + (44\alpha^2 + 72\alpha + 3)\theta^4 \\ +8(11\alpha+3)\theta^3\alpha + 4(8\alpha+11)\theta^2\alpha^2 + 32\theta\alpha^3 + 8\alpha^4 \end{Bmatrix}}{\left\{\theta^4 + 2\theta^3 + (4\alpha+1)\theta^2 + 4\theta\alpha + 2\alpha^2\right\}^2}$$

$$\gamma = \frac{\sigma^2}{\mu_1'} = \frac{\theta^4 + 2\theta^3 + (4\alpha+1)\theta^2 + 4\theta\alpha + 2\alpha^2}{\theta\left(\theta^2+\theta+\alpha\right)\left(\theta^2+\theta+2\alpha\right)}.$$

Graphs of C.V, $\sqrt{\beta_1}$, β_2 and γ of QGD for varying values of the parameters θ and α have been presented in figure 3.The C.V is monotonically decreasing for increasing values of the parameters θ and α but for increasing value of the parameter θ, the C.V shifts upward. The nature of coefficient of skewness (C.S) is also similar to the nature of C.V. The coefficient of kurtosis (C.K) is also monotonically decreasing for increasing values of the parameters θ and α but for $\theta = 0.5$ and, $\alpha \geq 2$, the C.K become constant. QGD is over-dispersed $(m < s^2)$, equi-dispersed $(m = s^2)$ and under-dispersed $(m > s^2)$ for $(q < 1, a^3 \; 0)$, $(q = 1, a^3 \; 0)$ and $(q > 1, a^3 \; 0)$ respectively, which is obvious from the graphs of index of dispersion (I.D) in figure 3.



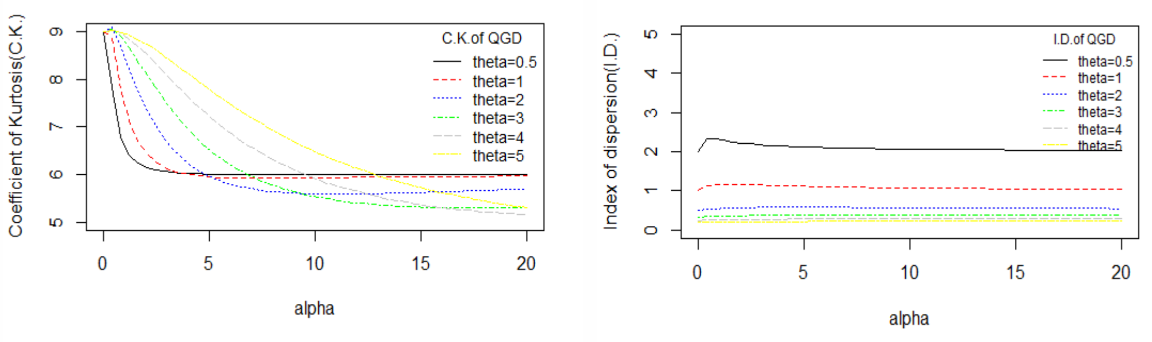


Figure 3. Graphs of C.V, $\sqrt{\beta_1}$, β_2 and γ of QGD for varying values of parameters θ and α

4. Reliability properties

Suppose X is a continuous random variable with pdf $f(x)$ and cdf $F(x)$. The hazard rate function (also known as the failure rate function) $h(x)$ and the mean residual life function $m(x)$ of X are respectively defined as

$$h(x)=\lim_{\Delta x\to 0}\frac{P\left(X<x+\Delta x \mid X>x\right)}{\Delta x}=\frac{f(x)}{1-F(x)} \tag{4.1}$$

and $m(x)=E\left[X-x \mid X>x\right]=\dfrac{1}{1-F(x)}\displaystyle\int_x^{\infty}\left[1-F(t)\right]\,dt$ (4.2)

Thus corresponding $h(x)$ and $m(x)$ of QGD are thus obtained as

$$h(x)=\frac{f_2(x;\theta,\alpha)}{1-F_2(x;\theta,\alpha)}=\frac{\theta^2\left(\alpha x+\theta+1\right)}{\alpha\theta x+\theta^2+\theta+\alpha} \tag{4.3}$$

and $m(x)=\dfrac{1}{\left(\alpha\theta x+\theta^2+\theta+\alpha\right)e^{-\theta x}}\displaystyle\int_x^{\infty}\left(\alpha\theta t+\theta^2+\theta+\alpha\right)e^{-\theta t}dt$

$$=\frac{\alpha\theta x+\theta^2+\theta+2\alpha}{\theta\left(\alpha\theta x+\theta^2+\theta+\alpha\right)} \tag{4.4}$$

It can be easily verified that $h(0)=\dfrac{\theta^2(\theta+1)}{\theta^2+\theta+\alpha}=f(0)$ and $m(0)=\dfrac{\theta^2+\theta+2\alpha}{\theta\left(\theta^2+\theta+\alpha\right)}=\mu_1'$.

The nature and behavior of $h(x)$ and $m(x)$ of QGD for varying values of parameters θ and α have been shown graphically in figures 4 and 5. For fixed values of α

and increasing values of ***q***, the $h(x)$ is shifting upward with very minor decrease /increase , whereas for fixed ***q*** and increasing ***a***, $h(x)$ is monotonically increasing.

The graphs of $m(x)$ are monotonically increasing for increasing values of ***q*** and ***a***, which is obvious from fig. 4.

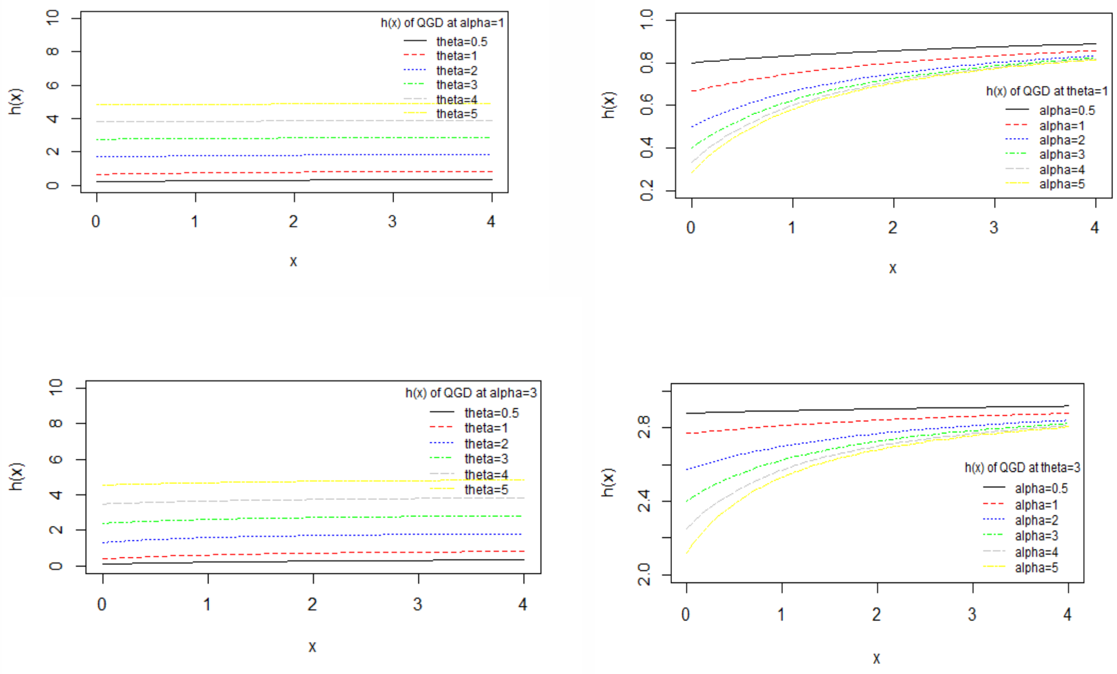


Figure 3. Graphs of $h(x)$ of QGD for varying values of parameters θ and α

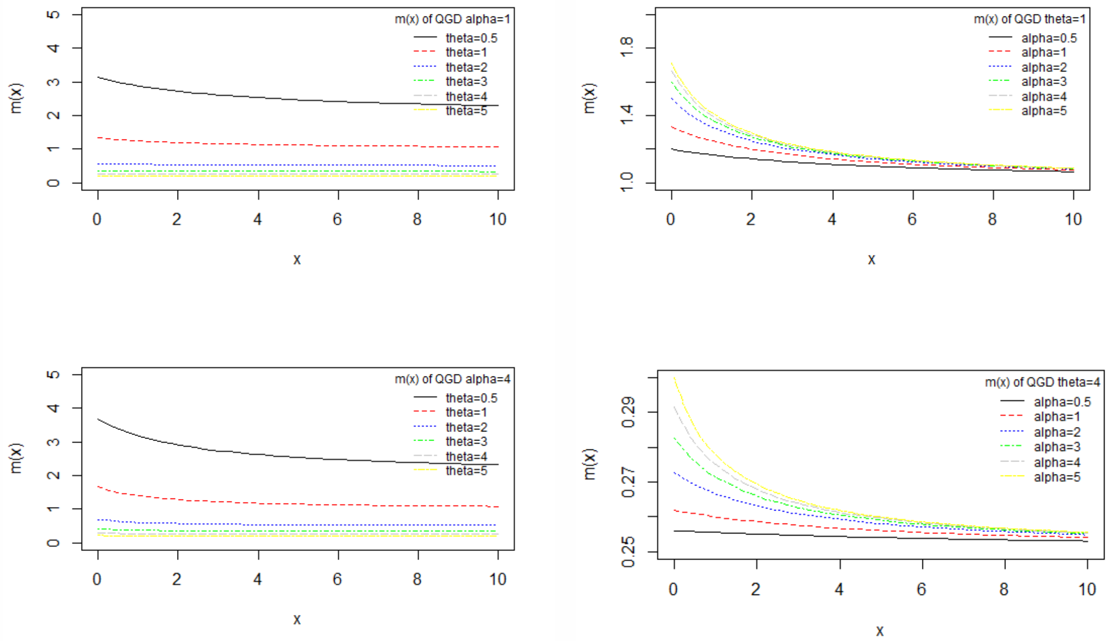


Figure 4. Graphs of $m(x)$ of QGD for varying values of parameters θ and α

5. Stochastic ordering

Stochastic ordering of positive continuous random variables is an important tool for judging their comparative behavior. A random variable X is said to be smaller than a random variable Y in the

(i) stochastic order $(X \le_{st} Y)$ if $F_X(x) \ge F_Y(x)$ for all x

(ii) hazard rate order $(X \le_{hr} Y)$ if $h_X(x) \ge h_Y(x)$ for all x

(iii) mean residual life order $(X \le_{mrl} Y)$ if $m_X(x) \le m_Y(x)$ for all x

(iv) likelihood ratio order $(X \le_{lr} Y)$ if $\dfrac{f_X(x)}{f_Y(x)}$ decreases in x.

The following results due to Shaked and Shanthikumar (1994) are well known for establishing stochastic ordering of distributions

$$\begin{array}{c} X \le_{lr} Y \Rightarrow X \le_{hr} Y \Rightarrow X \le_{mrl} Y \\ \Downarrow \\ X \le_{st} Y \end{array}$$

The QGD is ordered with respect to the strongest 'likelihood ratio ordering' as shown in the following theorem:

Theorem: Let $X \sim$ QGD(θ_1, α_1) and $Y \sim$ QGD(θ_2, α_2). If $\alpha_1 = \alpha_2$ and $\theta_1 > \theta_2$ (or $\theta_1 = \theta_2$ and $\alpha_1 < \alpha_2$), then $X \le_{lr} Y$ and hence $X \le_{hr} Y$, $X \le_{mrl} Y$ and $X \le_{st} Y$.

Proof: We have

$$\frac{f_X(x;\theta_1,\alpha_1)}{f_Y(x;\theta_2,\alpha_2)} = \frac{\theta_1^2\left(\theta_2^2+\theta_2+\alpha_2\right)}{\theta_2^2\left(\theta_1^2+\theta_1+\alpha_1\right)}\left(\frac{1+\theta_1+\alpha_1 x}{1+\theta_2+\alpha_2 x}\right)e^{-(\theta_1-\theta_2)x} \quad ; x > 0$$

Now

$$\ln\frac{f_X(x;\theta_1,\alpha_1)}{f_Y(x;\theta_2,\alpha_2)} = \ln\left[\frac{\theta_1^2\left(\theta_2^2+\theta_2+\alpha_2\right)}{\theta_2^2\left(\theta_1^2+\theta_1+\alpha_1\right)}\right] + \ln\left(\frac{1+\theta_1+\alpha_1 x}{1+\theta_2+\alpha_2 x}\right) - (\theta_1-\theta_2)x.$$

This gives

$$\frac{d}{dx}\left\{\ln\frac{f_X(x;\theta_1,\alpha_1)}{f_Y(x;\theta_2,\alpha_2)}\right\} = \frac{(\alpha_1-\alpha_2)+(\alpha_1\theta_2-\alpha_2\theta_1)}{(1+\theta_1+\alpha_1 x)(1+\theta_2+\alpha_2 x)} - (\theta_1-\theta_2).$$

Thus if $\alpha_1 = \alpha_2$ and $\theta_1 > \theta_2$ or $\theta_1 = \theta_2$ and $\alpha_1 < \alpha_2$, $\dfrac{d}{dx}\ln\dfrac{f_X(x;\theta_1,\alpha_1)}{f_Y(x;\theta_2,\alpha_2)} < 0$. This means that $X \le_{lr} Y$ and hence $X \le_{hr} Y$, $X \le_{mrl} Y$ and $X \le_{st} Y$.

6. Mean deviations

The amount of scatter in a population is measured to some extent by the totality of deviations usually from mean and median. These are known as the mean deviation about the mean and the mean deviation about the median defined by

$\delta_1(X)=\int_0^\infty |x-\mu|\, f(x)\,dx$ and $\delta_2(X)=\int_0^\infty |x-M|\, f(x)\,dx$, respectively, where $\mu=E(X)$ and $M=\text{Median}(X)$. The measures $\delta_1(X)$ and $\delta_2(X)$ can be calculated using the following simplified relationships

$$\delta_1(X)=\int_0^\mu (\mu-x)f(x)\,dx+\int_\mu^\infty (x-\mu)f(x)\,dx$$

$$=\mu F(\mu)-\int_0^\mu x\, f(x)\,dx-\mu\left[1-F(\mu)\right]+\int_\mu^\infty x\, f(x)\,dx$$

$$=2\mu F(\mu)-2\mu+2\int_\mu^\infty x\, f(x)\,dx$$

$$=2\mu F(\mu)-2\int_0^\mu x\, f(x)\,dx \qquad (6.1)$$

and

$$\delta_2(X)=\int_0^M (M-x)f(x)\,dx+\int_M^\infty (x-M)f(x)\,dx$$

$$=M\,F(M)-\int_0^M x\, f(x)\,dx-M\left[1-F(M)\right]+\int_M^\infty x\, f(x)\,dx$$

$$=-\mu+2\int_M^\infty x\, f(x)\,dx$$

$$=\mu-2\int_0^M x\, f(x)\,dx \qquad (6.2)$$

Using pdf of QGD (2.1) and expression for the mean of QGD, we get

$$\int_0^\mu x\, f_2(x;\theta,\alpha)\,dx=\mu-\frac{\left\{\alpha\,\theta^2\mu^2+\theta\left(\theta^2+\theta+2\alpha\right)\mu+\left(\theta^2+\theta+2\alpha\right)\right\}e^{-\theta\mu}}{\theta\left(\theta^2+\theta+\alpha\right)} \qquad (6.3)$$

$$\int_0^M x\, f_2(x;\theta,\alpha)\,dx=\mu-\frac{\left\{\alpha\,\theta^2M^2+\theta\left(\theta^2+\theta+2\alpha\right)M+\left(\theta^2+\theta+2\alpha\right)\right\}e^{-\theta M}}{\theta\left(\theta^2+\theta+\alpha\right)} \qquad (6.4)$$

Using expressions from (6.1), (6.2), (6.3), and (6.4), the mean deviation about mean, $\delta_1(X)$ and the mean deviation about median, $\delta_2(X)$ of QGD are finally obtained as

$$\delta_1(X)=\frac{2\left\{\alpha\,\theta\,\mu+\left(\theta^2+\theta+2\alpha\right)\right\}e^{-\theta\mu}}{\theta\left(\theta^2+\theta+\alpha\right)} \qquad (6.5)$$

$$\delta_2(X)=\frac{2\left\{\alpha\theta^2M^2+\theta\left(\theta^2+\theta+2\alpha\right)M+\left(\theta^2+\theta+2\alpha\right)\right\}e^{-\theta M}}{\theta\left(\theta^2+\theta+\alpha\right)}-\mu \tag{6.6}$$

7. Bonferroni and Lorenz curves

The Bonferroni and Lorenz curves (Bonferroni (1930)) and Bonferroni and Gini indices have applications not only in economics to study income and poverty, but also in other fields like reliability, demography, insurance and medicine. The Bonferroni and Lorenz curves are defined as

$$B(p)=\frac{1}{p\mu}\int_0^q x\,f(x)dx=\frac{1}{p\mu}\left[\int_0^\infty x\,f(x)dx-\int_q^\infty x\,f(x)dx\right]=\frac{1}{p\mu}\left[\mu-\int_q^\infty x\,f(x)dx\right] \tag{7.1}$$

and $$L(p)=\frac{1}{\mu}\int_0^q x\,f(x)dx=\frac{1}{\mu}\left[\int_0^\infty x\,f(x)dx-\int_q^\infty x\,f(x)dx\right]=\frac{1}{\mu}\left[\mu-\int_q^\infty x\,f(x)dx\right] \tag{7.2}$$

respectively or equivalently

$$B(p)=\frac{1}{p\mu}\int_0^p F^{-1}(x)dx \tag{7.3}$$

and $$L(p)=\frac{1}{\mu}\int_0^p F^{-1}(x)dx \tag{7.4}$$

respectively, where $\mu=E(X)$ and $q=F^{-1}(p)$.

The Bonferroni and Gini indices are thus defined as

$$B=1-\int_0^1 B(p)dp \tag{7.5}$$

and $$G=1-2\int_0^1 L(p)dp \tag{7.6}$$

respectively.

Using pdf of QGD (2.1), we get

$$\int_q^\infty x\,f_2(x;\theta,\alpha)dx=\frac{\left\{\alpha\theta^2q^2+\theta\left(\theta^2+\theta+2\alpha\right)q+\left(\theta^2+\theta+2\alpha\right)\right\}e^{-\theta q}}{\theta\left(\theta^2+\theta+\alpha\right)} \tag{7.7}$$

Now using equation (7.7) in (7.1) and (7.2), we get

$$B(p)=\frac{1}{p}\left[1-\frac{\left\{\alpha\theta^2q^2+\theta\left(\theta^2+\theta+2\alpha\right)q+\left(\theta^2+\theta+2\alpha\right)\right\}e^{-\theta q}}{\theta^2+\theta+2\alpha}\right] \tag{7.8}$$

and

$$L(p)=1-\frac{\left\{\alpha\theta^2q^2+\theta\left(\theta^2+\theta+2\alpha\right)q+\left(\theta^2+\theta+2\alpha\right)\right\}e^{-\theta q}}{\theta^2+\theta+2\alpha} \tag{7.9}$$

Now using equations (7.8) and (7.9) in (7.5) and (7.6), the Bonferroni and Gini indices of QGD are thus obtained as

$$B=1-\frac{\left\{\alpha\theta^2q^2+\theta\left(\theta^2+\theta+2\alpha\right)q+\left(\theta^2+\theta+2\alpha\right)\right\}e^{-\theta q}}{\theta^2+\theta+2\alpha} \tag{7.10}$$

$$G=\frac{2\left\{\alpha\theta^2q^2+\theta\left(\theta^2+\theta+2\alpha\right)q+\left(\theta^2+\theta+2\alpha\right)\right\}e^{-\theta q}}{\theta^2+\theta+2\alpha}-1 \tag{7.11}$$

8. Order statistics and Renyi entropy measure

8.1. Distribution of Order Statistics

Let $X_1, X_2, ..., X_n$ be a random sample of size n from QGD (2.1). Let $X_{(1)} < X_{(2)} < ... < X_{(n)}$ denote the corresponding order statistics. The pdf and the cdf of the k th order statistic, say $Y = X_{(k)}$ are given by

$$f_Y(y)=\frac{n!}{(k-1)!(n-k)!}F^{k-1}(y)\{1-F(y)\}^{n-k}f(y)$$

$$=\frac{n!}{(k-1)!(n-k)!}\sum_{l=0}^{n-k}\binom{n-k}{l}(-1)^l F^{k+l-1}(y)f(y)$$

and

$$F_Y(y)=\sum_{j=k}^{n}\binom{n}{j}F^j(y)\{1-F(y)\}^{n-j}$$

$$=\sum_{j=k}^{n}\sum_{l=0}^{n-j}\binom{n}{j}\binom{n-j}{l}(-1)^l F^{j+l}(y),$$

respectively, for $k=1,2,3,...,n$.

Thus, the pdf and the cdf of k th order statistic of QGD are thus obtained as

$$f_Y(y)=\frac{n!\theta^2(1+\theta+\alpha x)e^{-\theta x}}{\left(\theta^2+\theta+\alpha\right)(k-1)!(n-k)!}\times\sum_{l=0}^{n-k}\binom{n-k}{l}(-1)^l$$
$$\times\left[1-\frac{(\theta^2+\theta+\alpha)+\alpha\theta x}{\theta^2+\theta+\alpha}e^{-\theta x}\right]^{k+l-1}$$

and

$$F_Y(y)=\sum_{j=k}^{n}\sum_{l=0}^{n-j}\binom{n}{j}\binom{n-j}{l}(-1)^l\left[1-\frac{(\theta^2+\theta+\alpha)+\alpha\theta x}{\theta^2+\theta+\alpha}\right]^{j+l}$$

8.2. Renyi Entropy Measure

An entropy of a random variable X is a measure of variation of uncertainty. A popular entropy measure is Renyi entropy (1961). If X is a continuous random variable having pdf $f(.)$, then Renyi entropy is defined as

$$T_R(\gamma)=\frac{1}{1-\gamma}\log\left\{\int f^{\gamma}(x)\,dx\right\}$$

where $\gamma>0$ and $\gamma\neq1$.

Thus, the Renyi entropy of QGD can be obtained as

$$T_R(\gamma)=\frac{1}{1-\gamma}\log\left[\int_0^\infty \frac{\theta^{2\gamma}}{\left(\theta^2+\theta+\alpha\right)^{\gamma}}(1+\theta+\alpha x)^{\gamma}e^{-\theta\gamma x}dx\right]$$

$$=\frac{1}{1-\gamma}\log\left[\int_0^\infty \frac{\theta^{2\gamma}(1+\theta)^{\gamma}}{\left(\theta^2+\theta+\alpha\right)^{\gamma}}\left(1+\frac{\alpha x}{1+\theta}\right)^{\gamma}e^{-\theta\gamma x}dx\right]$$

$$=\frac{1}{1-\gamma}\log\left[\int_0^\infty \frac{\theta^{2\gamma}(1+\theta)^{\gamma}}{\left(\theta^2+\theta+\alpha\right)^{\gamma}}\sum_{j=0}^{\infty}\binom{\gamma}{j}\left(\frac{\alpha x}{1+\theta}\right)^{j}e^{-\theta\gamma x}dx\right]$$

$$=\frac{1}{1-\gamma}\log\left[\sum_{j=0}^{\infty}\binom{\gamma}{j}\frac{\alpha^{j}\theta^{2\gamma}(1+\theta)^{\gamma-j}}{\left(\theta^2+\theta+\alpha\right)^{\gamma}}\int_0^\infty e^{-\theta\gamma x}x^{j+1-1}dx\right]$$

$$=\frac{1}{1-\gamma}\log\left[\sum_{j=0}^{\infty}\binom{\gamma}{j}\frac{\alpha^{j}\theta^{2\gamma}(1+\theta)^{\gamma-j}}{\left(\theta^2+\theta+\alpha\right)^{\gamma}}\frac{\Gamma(j+1)}{(\theta\gamma)^{j+1}}\right]$$

$$=\frac{1}{1-\gamma}\log\left[\sum_{j=0}^{\infty}\binom{\gamma}{j}\frac{\alpha^{\gamma}\theta^{2\gamma-j-1}(1+\theta)^{\gamma-j}}{\left(\theta^2+\theta+\alpha\right)^{\gamma}}\frac{\Gamma(j+1)}{(\gamma)^{j+1}}\right]$$

9. Stress-strength reliability

The stress- strength reliability describes the life of a component which has random strength X that is subjected to a random stress Y. When the stress applied to it exceeds the strength, the component fails instantly and the component will function satisfactorily till

$X > Y$. Therefore, $R = P(Y < X)$ is a measure of component reliability and in statistical literature it is known as stress-strength parameter. It has wide applications in almost all areas of knowledge especially in engineering such as structures, deterioration of rocket motors, static fatigue of ceramic components, aging of concrete pressure vessels etc.

Let X and Y be independent strength and stress random variables having QGD (2.1) with parameter (θ_1, α_1) and (θ_2, α_2) respectively. Then the stress-strength reliability R of QGD (2.1) can be obtained as

$$R = P(Y < X) = \int_0^\infty P(Y < X \mid X = x) f_X(x)\,dx$$

$$= \int_0^\infty f_2(x;\theta_1,\alpha_1)\, F_2(x;\theta_2,\alpha_2)\,dx$$

$$= 1 - \frac{\theta_1^2 \begin{bmatrix} (\theta_1+\theta_2)^2(\theta_1+1)\left(\theta_2^2+\theta_2+\alpha_2\right)+2(\alpha_1\alpha_2\theta_2) \\ +(\theta_1+\theta_2)\left\{\alpha_2\theta_2(\theta_1+1)+\alpha_1\left(\theta_2^2+\theta_2+\alpha_2\right)\right\} \end{bmatrix}}{\left(\theta_1^2+\theta_1+\alpha_1\right)\left(\theta_2^2+\theta_2+\alpha_2\right)(\theta_1+\theta_2)^3}.$$

It can be easily verified that at $\alpha_1 = \theta_1$ and $\alpha_2 = \theta_2$, the above expression reduces to R of Garima distribution.

10. Estimation of parameters

10.1. Method of Moment Estimates (MOME)

Since QGD have two parameters to be estimated, the first two moments about origin are required to get method of moment estimates. Equating the sample mean to the corresponding population mean, we get

$$\overline{x} = \frac{\theta^2+\theta+2\alpha}{\theta\left(\theta^2+\theta+\alpha\right)} = \frac{1}{\theta} + \frac{\alpha}{\theta^2+\theta+\alpha}$$

This gives

$$\theta^2+\theta+\alpha = \frac{\alpha}{\theta\overline{x}-1} \qquad (9.1.1)$$

Again equating the second sample moment to corresponding second population moment, we get

$$m_2' = \frac{2\left(\theta^2+\theta+3\alpha\right)}{\theta^2\left(\theta^2+\theta+\alpha\right)} = \frac{2}{\theta^2} + \frac{4\alpha}{\theta^2\left(\theta^2+\theta+\alpha\right)}$$

This gives

$$\theta^2+\theta+\alpha = \frac{4\alpha}{m_2'\theta^2-2} \qquad (9.1.2)$$

Using equations (9.1.1) and (9.1.2), we get a quadratic equation in θ as

$$m_2'\theta^2 - 4\overline{x}\theta + 2 = 0.$$

This gives the MOME estimate $\tilde{\theta}$ of θ as

$$\tilde{\theta} = \frac{2\overline{x} + \sqrt{4\overline{x}^2 - 2m_2'}}{m_2'} \quad ; m_2' < 2\overline{x}^2$$

Substituting the value of $\tilde{\theta}$ in equation (9.1.1) , the MOME $\tilde{\alpha}$ of α is given by

$$\tilde{\alpha} = \frac{\theta(\theta+1)(\theta\,\overline{x}-1)}{2-\theta\overline{x}}$$

10.2. Maximum Likelihood Estimates (MLE)

Let $(x_1, x_2, \ldots, x_n)$ be a random sample of size n from QGD (2.1)). The likelihood function, L of QGD is given by

$$L = \left(\frac{\theta^2}{\theta^2+\theta+\alpha}\right)^n \prod_{i=1}^{n}(1+\theta+\alpha\,x_i)e^{-n\,\theta\,\overline{x}}$$

The natural log likelihood function is thus obtained as

$$\ln L = n\ln\left(\frac{\theta^2}{\theta^2+\theta+\alpha}\right) + \sum_{i=1}^{n}\ln(1+\theta+\alpha\,x_i) - n\,\theta\,\overline{x}$$

The maximum likelihood estimates (MLE) $(\hat{\theta}, \hat{\alpha})$ of (θ, α) are then the solutions of the following non-linear equations

$$\frac{\partial \ln L}{\partial \theta} = \frac{2n}{\theta} - \frac{n(2\theta+1)}{\theta^2+\theta+\alpha} - n\,\overline{x} + \sum_{i=1}^{n}\frac{1}{1+\theta+\alpha x_i} = 0$$

$$\frac{\partial \ln L}{\partial \alpha} = \frac{-n}{\theta^2+\theta+\alpha} + \sum_{i=1}^{n}\frac{x_i}{1+\theta+\alpha x_i} = 0$$

where $\overline{x}$ is the sample mean.

These two natural log likelihood equations do not seem to be solved directly because they are not in closed forms. However, the Fisher's scoring method can be applied to solve these equations. For, we have

$$\frac{\partial^2 \ln L}{\partial \theta^2} = -\frac{2n}{\theta^2} + \frac{n(2\theta^2+2\theta-2\alpha+1)\alpha^2}{(\theta^2+\theta+\alpha)^2} - \sum_{i=1}^{n}\frac{1}{(1+\theta+\alpha\,x_i)^2}$$

$$\frac{\partial^2 \ln L}{\partial \theta\,\partial \alpha} = \frac{n(2\theta+1)}{(\theta^2+\theta+\alpha)^2} - \sum_{i=1}^{n}\frac{x_i}{(1+\theta+\alpha\,x_i)^2}$$

$$\frac{\partial^2 \ln L}{\partial \alpha^2} = \frac{n}{(\theta^2+\theta+\alpha)^2} - \sum_{i=1}^{n}\frac{x_i^{\,2}}{(1+\theta+\alpha\,x_i)^2}$$

The solution of following equations gives MLE's $(\hat{\theta}, \hat{\alpha})$ of (θ, α) of QGD

$$\begin{bmatrix} \dfrac{\partial^2 \ln L}{\partial \theta^2} & \dfrac{\partial^2 \ln L}{\partial \theta \partial \alpha} \\ \dfrac{\partial^2 \ln L}{\partial \theta \partial \alpha} & \dfrac{\partial^2 \ln L}{\partial \alpha^2} \end{bmatrix}_{\substack{\hat{\theta}=\theta_0 \\ \hat{\alpha}=\alpha_0}} \begin{bmatrix} \hat{\theta}-\theta_0 \\ \hat{\alpha}-\alpha_0 \end{bmatrix} = \begin{bmatrix} \dfrac{\partial \ln L}{\partial \theta} \\ \dfrac{\partial \ln L}{\partial \alpha} \end{bmatrix}_{\substack{\hat{\theta}=\theta_0 \\ \hat{\alpha}=\alpha_0}}$$

where θ_0 and α_0 are the initial values of θ and α as given by the MOME of QGD. These equations are solved iteratively till sufficiently close values of $\hat{\theta}$ and $\hat{\alpha}$ are obtained.

11. Data analysis

In this section, the goodness of fit of QGD has been discussed with a real lifetime dataset from engineering and the fit has been compared with one parameter and two-parameter lifetime distributions. The following dataset represents the failure times (in minutes) for a sample of 15 electronic components in an accelerated life test, Lawless (2003)

1.4 5.1 6.3 10.8 12.1 18.5 19.7 22.2 23.0 30.6 37.3 46.3 53.9 59.8 66.2

In order to compare the considered distributions, values of $-2\ln L$, AIC(Akaike Information Criterion) and K-S Statistic (Kolmogorov-Smirnov Statistic) and p-value for the dataset have been computed and presented in table 2. The AIC and K-S Statistic are defined as follow:

$AIC = -2\ln L + 2k$ and $K\text{-}S = \operatorname{Sup}_x \left| F_n(x) - F_0(x) \right|$, where k =number of parameters, n = sample size, $F_n(x)$ is the empirical distribution function and $F_0(x)$ is the theoretical cumulative distribution function.. The best distribution corresponds to the lower values of $-2\ln L$, AIC, K-S statistic and higher p-value.

The pdf and the cdf of the fitted distributions have been given in table 1. Recall that the quasi Shanker distribution (QSD) has been introduced by Shanker & Shukla (2017) and the exponentiated exponential distribution (EED) has been introduced by Gupta & Kundu (1999). The Lindley distribution has been introduced by Lindley (1958) and its detailed study has been done by Ghitany *et al* (2008).

Table 1. The pdf and the cdf of the fitted distributions

Models	p.d.f.	c.d.f.
QSD	$f(x;\theta,\alpha) = \left(\dfrac{\theta^3}{\theta^3+\theta+2\alpha}\right) \times \left(\theta + x + \alpha x^2\right) e^{-\theta x}$	$F(x;\theta,\alpha) = 1 - \left[1 + \dfrac{\left\{\begin{matrix}\alpha\theta^2 x^2 \\ +\theta x(\theta+2\alpha)\end{matrix}\right\}}{\theta^3+\theta+2\alpha}\right] e^{-\theta x}$
Weibull	$f(x;\theta,\alpha) = \theta\alpha x^{\alpha-1} e^{-\theta x^{\alpha}}$	$F(x;\theta,\alpha) = 1 - e^{-\theta x^{\alpha}}$

Gamma	$f(x;\theta,\alpha)=\dfrac{\theta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\theta x}$	$F(x;\theta,\alpha)=1-\dfrac{\Gamma(\alpha,\theta x)}{\Gamma(\alpha)}$
Lognormal	$f(x;\theta,\alpha)=\dfrac{1}{\sqrt{2\pi}\alpha x}e^{-\frac{1}{2}\left(\frac{\log x-\theta}{\alpha}\right)^2}$	$F(x;\theta,\alpha)=\phi\left(\dfrac{\log x-\theta}{\alpha}\right)$
EED	$f(x;\theta,\alpha)=\alpha\,\theta\left(1-e^{-\theta x}\right)^{\alpha-1}e^{-\theta x}$	$F(x;\theta,\alpha)=\left(1-e^{-\theta x}\right)^{\alpha}$
Lindley	$f(x;\theta)=\dfrac{\theta^2}{\theta+1}(1+x)e^{-\theta x}$	$F(x;\theta)=1-\left[\dfrac{\theta+1+\theta x}{\theta+1}\right]e^{-\theta x}$
Exponential	$f(x;\theta)=\theta e^{-\theta x}$	$F(x;\theta)=1-e^{-\theta x}$

Table 2. MLE's, $-2\ln L$, Standard Error (S.E), AIC, K-S Statistics and p-value of the fitted distributions of dataset.

Distributions	ML Estimates	S.E	$-2\ln L$	AIC	K-S	p-value
QGD	$\hat{\theta}=0.06225$	0.01721	128.21	132.21	0.095	0.997
	$\hat{\alpha}=0.16577$	0.32075				
QSD	$\hat{\theta}=0.07389$	0.04131	129.37	133.37	0.121	0.961
	$\hat{\alpha}=0.00147$	0.04401				
Gamma	$\hat{\theta}=0.05236$	0.02067	128.37	132.37	0.102	0.992
	$\hat{\alpha}=1.44219$	0.47771				
Weibull	$\hat{\theta}=0.01190$	0.01124	128.04	132.04	0.098	0.995
	$\hat{\alpha}=1.30586$	0.24925				
Lognormal	$\hat{\theta}=2.93059$	0.26472	131.23	135.23	0.161	0.951
	$\hat{\alpha}=1.02527$	0.18718				
EED	$\hat{\theta}=0.04529$	0.01372	128.47	132.47	0.108	0.986
	$\hat{\alpha}=1.44347$	0.51301				
Garima	$\hat{\theta}=0.05462$	0.01227	128.52	130.52	0.123	0.954
Lindley	$\hat{\theta}=0.07022$	0.01283	128.81	130.81	0.110	0.983
Exponential	$\hat{\theta}=0.03631$	0.00937	129.47	131.47	0.156	0.807

It can be easily seen from table 2 that the QGD gives better fit than one parameter exponential, Lindley and Garima distributions and two-parameter QSD, Gamma, Weibull lognormal and EED and hence it can be considered as an important distribution for modeling lifetime dataset over these distributions.

12. Concluding remarks

A two-parameter quasi Garima distribution (QGD), of which one parameter exponential distribution and Garima distribution introduced by Shanker (2016 c) are a particular cases, has been suggested and investigated. Its mathematical properties including moments, coefficient of variation, skewness, kurtosis, index of dispersion, hazard rate function, mean residual life function, stochastic ordering, mean deviations, Bonferroni and Lorenz curves, order statistics, Renyi entropy measure and stress-strength reliability have been discussed.

For estimating its parameters the method of moments and the method of maximum likelihood estimation have been discussed. Finally, a numerical example of real lifetime dataset has been presented to test the goodness of fit of QGD over one parameter exponential, Lindley and Garima distributions and two-parameter QSD, Gamma, Weibull lognormal and EED and the fit by QGD has been found to be quite satisfactory. Therefore, QGD can be recommended as an important two-parameter lifetime distribution for modeling lifetime data over these distributions.

References

1. Bonferroni, C.E. **Elementi di Statistica generale,** Seeber, Firenze, 1930
2. Ghitany, M.E., Atieh, B. & Nadarajah, S. **Lindley distribution and its Application,** Mathematics Computing and Simulation, Vol. 78, 2008, pp. 493–506.
3. Gupta, R.D. & Kundu, D. **Generalized Exponential Distributions,** Australian and New Zealand Journal of Statistics, Vol. 41, Issue 2, 1999, pp. 173–188.
4. Lawless, J.F. **Statistical models and methods for lifetime data,** John Wiley and Sons, New York, 2003
5. Lindley, D.V. **Fiducial distributions and Bayes' theorem,** Journal of the Royal Statistical Society, Series B, Vol. 20, 1958, pp. 102- 107.
6. Renyi, A. **On measures of entropy and information,** in "Proceedings of the 4th Berkeley symposium on Mathematical Statistics and Probability", Berkeley, University of California press, Vol. 1, 1961, pp. 547 – 561.
7. Shaked, M. & Shanthikumar, J.G. **Stochastic Orders and Their Applications,** Academic Press, New York, 1994
8. Shanker, R. **Shanker Distribution and Its Applications,** International Journal of Statistics and Applications, *Vol.* 5, Issue 6, 2015a, pp. 338–348.
9. Shanker, R. **Akash Distribution and Its Applications,** International Journal of Probability and Statistics, Vol. 4, Issue 3, 2015b, pp. 65–75.
10. Shanker, R. **Aradhana Distribution and Its Applications,** International Journal of Statistics and Applications, Vol. 6, Issue 1, 2016a, pp. 23–34.
11. Shanker, R. **Sujatha Distribution and Its Applications,** Statistics in Transition - New series, Vol. 17, Issue 3, 2016b, pp. 1–20.
12. Shanker, R. **Garima distribution and Its application to model behavioral science data,** Biometrics & Biostatistics International Journal, Vol. 4, Issue 7, 2016c, pp. 1–9.
13. Shanker, R. **The discrete Poisson-Garima distribution,** Biometrics & Biostatistics International Journal, Vol. 5, Issue 2, 2017, pp. 1–7.
14. Shanker, R. & Shukla, K.K. **A Quasi Shanker Distribution and Its Applications,** Biometrics & Biostatistics International Journal, Vol. 6, Issue 1, 2017, pp. 1–10.

ANALYSIS OF THE RISK FACTORS OF DENTAL CARIES

Daniela CALOTA
PhD Candidate
University of Medicine and Pharmacy "Carol Davila", Bucharest, Romania

Abstract
According to World Health Organization data and statistics, dental caries is the leading oral disease by the frequency of occurrence. Tooth decay affects the people worldwide, becoming a major burden for the oral health. A large number of studies in the field have identified a series of risk factors that are responsible for causing dental caries. Therefore, in order to prevent this condition, we should put an emphasis on oral hygiene, healthy diet and food habits, frequency of visiting an oral health professional and relevant sources of information on oral health. For the study upon the risk factors concerning dental health, I applied a survey questionnaire on a sample of 514 adult people (over 18 years old). If we analyze the results, we can observe that 76.07 percent of the survey respondents change their toothbrush every three months, but 17.51 percent do not use additional hygiene practices. Regarding the dental visits, 44.16 percent of the survey respondents visit their dental practitioner only when they are in pain and less for a regular dental exam. Concerning the fruit and vegetables consumption, 31% of the survey respondents state that they have a poor consumption. To sum up, based on all the questionnaire data, there is a low interest of the population in their oral health, due to the lack of information on the risk factors. Despite the many oral health promotion campaigns, the results are not yet notable. It is significant that these campaigns should continue without interruption, differentiated by age groups, involving dentists, dental nurses, school teachers and the authorities.

Key words: dental caries; risk factors; patient behavior; quantitative analysis

1. Risk factors of dental caries

The tooth decay process is irreversible and has a long evolution, so it is often neglected by the patients. At the World Congress of Preventive Dentistry, The World Health Organization has placed the dental caries as the most prevalent among the oral diseases, according to the data presented, thus becoming a national and global problem because of the general health state of the population, as the dental caries increases the risk of developing digestive, cardiac, rheumatic or renal diseases by maintaining the infectious outbreak. At the same time, the aesthetic aspect of the patient is affected, along with the change in the social behaviour, the patient becoming more withdrawn.

According to a study of the National Institute of Public Health carried out through the National Centre for Health Evaluation and Promotion of Health (CNEPSS) in 2012, on a sample of 2851 children and adolescents, it was found that, at national level, the prevalence of dental caries was 3.39 caries per child.

The county with the highest average number of dental caries was Bistrița Năsăud with 5.07 caries per participant, and the minimum average number is in Galaţi County (1.23 caries per participant). Regarding the conditions for the dental caries incidence, numerous studies have shown that the tooth decay process is not present in the teeth that does not come into contact with the oral environment. From this we conclude that the risk of carious lesions occurs with the eruption of temporary teeth. With the onset of dental eruption, several contributing factors arise. Among the most important factors included are:

- the mineralized state of the dental structure-in case of the lack of calcium and mineral salts, the enamel has a low resistance to the aggressiveness of the microorganisms;
- the amount of saliva that promotes self-cleaning–the salivary flow removes interdental food remnants. When the amount of saliva decreases, these remnantsare maintained and ferment, thus creating a pH acid that affects the dental enamel;
- the bacterial plaque formation by the modification of the local pH - bacteria located on the dental surface modify through their metabolic activity the oral pH into an acid that affects the dental structures;
- the shape of the teeth and their position on the arcade - it was clinically observed that the lateral teeth (premolars, molars) can exhibit deep grooves on the occlusal surface, retaining food and creating difficulties in the self-cleansing process;
- eating habits that involve excessive consumption of carbohydrates and carbonated beverages to the detriment of dairy, fruit and vegetables. Excess carbohydrates (sugars) is an important factor in the adherence of food to the dental surface and also a nourishing substrate for acidogenic bacteria. Instead, dairy and fruit provide an intake of calcium, mineralsand vitamins essential for the resistance of dental tissues.

These factors may be associated with general disorders such as endocrine, thyroid, sexual or somatotropic disorders that can alter the moment of dental eruption, the mineralization of dental tissue, and saliva viscosity (salivary gland inflammation, salivary calculi, radiotherapy).

The carious process affects the entire population of the planet, so it has become a major problem in oro-dental health, to which we have to pay due attention.

2. Results

In the study based on the questionnaires we proposed to analyse the degree of exposure of the population to risk factors for dental health, as well as the degree of health education on oral hygiene. Thus, we conducted a prospective cohort analytical study on a sample of 514 adult adults aged over 18 years. The respondents are students of the post-secondary health schools and the questionnaire was distributed, with their consent, including a consent for data processing. This questionnaire was distributed between September 2016 and September 2017 and included questions about dental hygiene, the frequency of visits to a dentist, eating habits and sources of information on dental hygiene.

Of the 514 trainees who participated in this research, 415, representing 76.14% of the respondents, came from the urban area and only 99 trainees declared that they live in rural areas, representing 23.86%.

Table 1. Frequency of dental brushing among participants

Frequency of dental brushing	Number	%
Rare	**10**	**1**
One/day	**67**	**14**
Twice/day	**493**	**58**
After every meal	**343**	**27**

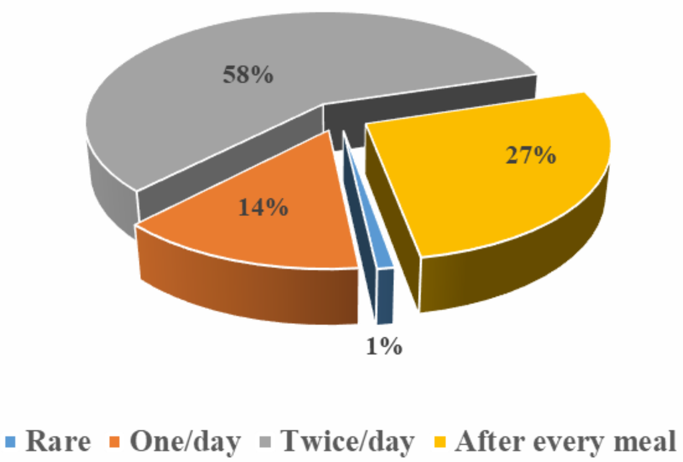


Figure 1. Frequency of dental brushing among participants

Table 2. Type of toothbrush

Type of toothbrush	Number	%
Hard	64	6
Medium	623	52
Soft	60	20
Electric	320	22

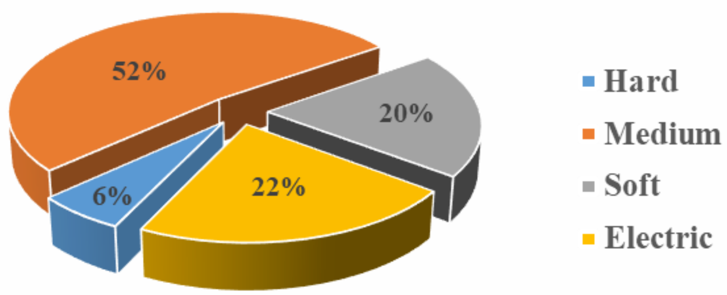


Figure 2. Type of toothbrush

Table 3. Toothbrush replacement

Time interval	Number	%
3 months	**391**	**76,07**
6 months	**99**	**19,26**
12 month	**24**	**4,67**

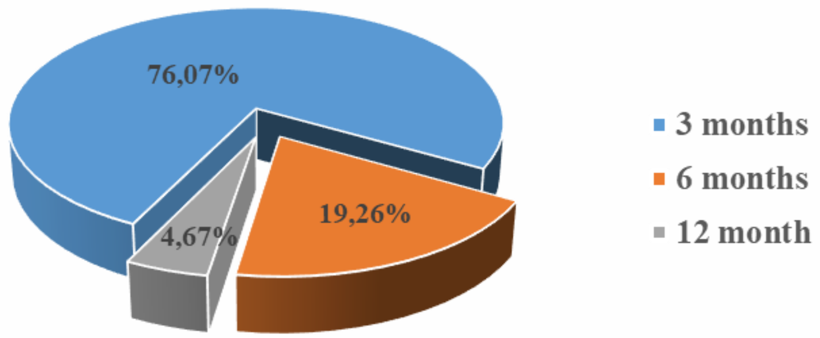


Figure 3. Toothbrush replacement

By analysing these charts, we can note that although 58.17% of the respondents perform dental brushing both in the morning and in the evening, a percentage of 14.98% brush their teeth once a day or less than once a day. The importance of dental brushing is generated by providing 93.77% of the individual's health and less emphasizing beauty or aesthetics. A fairly high percentage of the interviewees, namely 83.05%, use a correct choice of toothbrush, medium bristled brush and electric brush, respectively.

While 76% of respondents replace the toothbrush every 3 months, there are still some who replace their toothbrush every 6 months or even once a year. This demonstrates an insufficient knowledge of maintaining good dental hygiene and increased risk of tooth decay, due to the accumulation of bacterial plaque.

Table 4. Use of other hygiene tools

Other means used	**Number**	**%**
Dental floss	156	30,35
Mouthwash	104	20,23
Dental scaling	5	0,97
Mouthwash and dental floss	164	31,91
Did not respond	85	16,54

Table 5. Information sources in dental hygiene

Information sources	**Number**	**%**
Advertising materials	158	30,74
Internet	133	25,88
Dentist	223	43,38

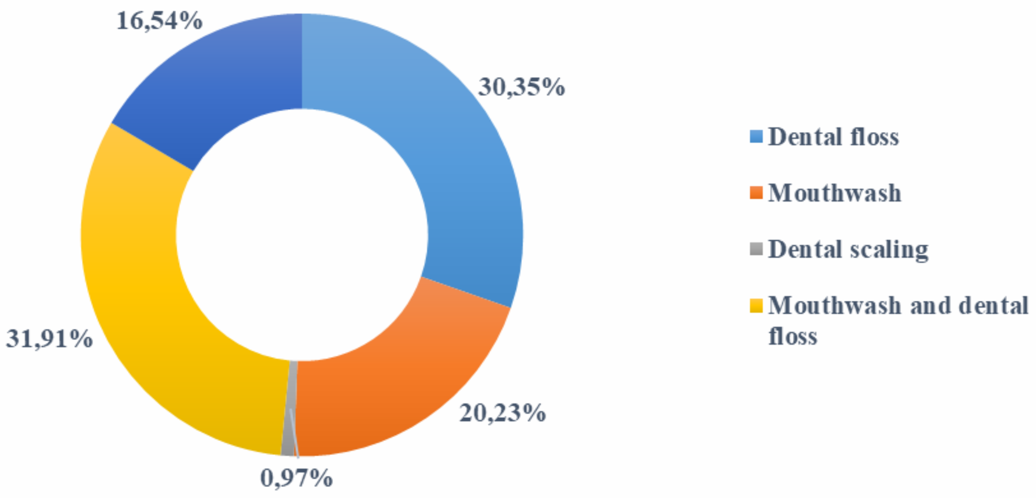


Figure 4. Use of other hygiene tools

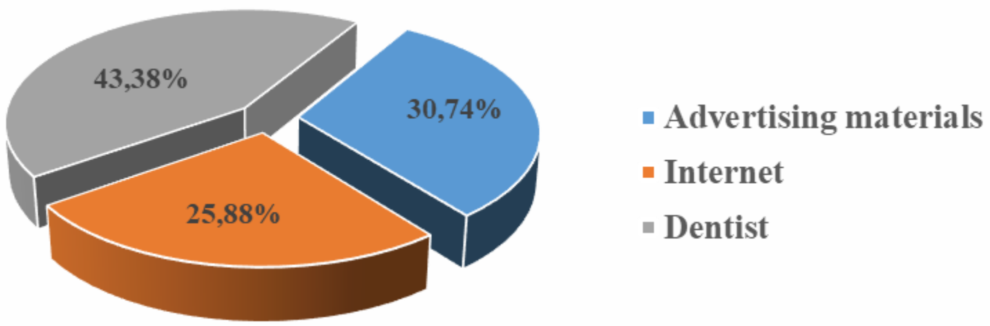
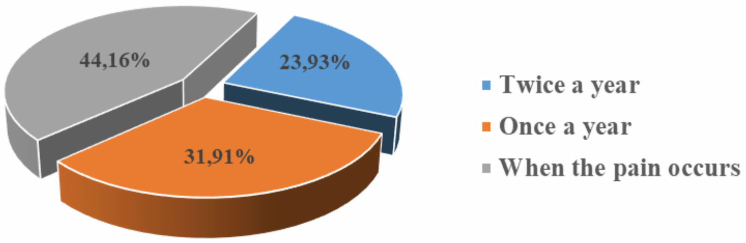


Figure 5. Information sources in dental hygiene

The centralized data in these two tables have shown that only 31.91% of the respondents use dental floss and mouthwash as additional oral hygiene tools, the others usingtheoral hygiene means partially or not at all, limiting their dental care only to toothbrush and toothpaste. One of these causes is also the fact that only 43.38% of the respondents are informed by the dentist about the means and modalities of maintaining their oral hygiene, the rest confining their information to the documentation provided by the Internet and various advertising materials. Most of the advertising materials regarding health, in general and oral health, in particular, are usually of a purely commercial nature and do not provide the best personalized information as tools, treatments and maintenance techniques.

Table 6. Frequency of dental examinations

Frequency of examinations	Number	%
Twice a year	123	23,93
Once a year	164	31,91
When the pain occurs	227	44,16



Graphic 6. Frequency of dental examinations

We can notice that only 23.93% of the respondents get their dental examination twice a year, correctly, while it is worrying that 44.16% of the respondents go for their examination only when the pain occurs, therefore only in emergency situations. This indicates poor individual concern for oral health and a low level of health education. Given that the individuals in the analysed sample are of adult age (22 years and over), the result shows that they did not benefit from proper health education in the childhood, which could also represent a risk of incorrect information on this issue for their future children.

Table 7. Consumption of carbonated soft drinks

Consumption	Number	%
Do not consume	70	13,62
Daily/ very often	196	38,13
Occasionally	248	48,25

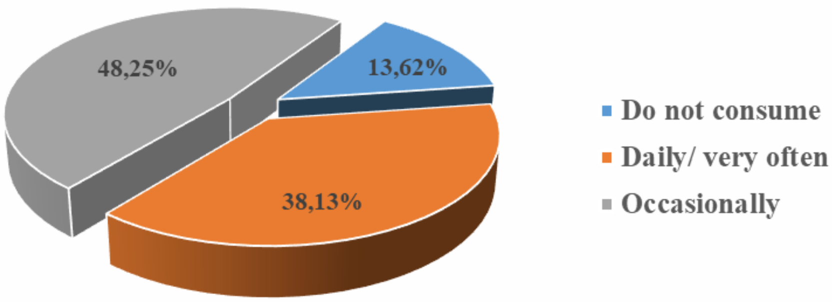


Figure 7. Consumption of carbonated soft drinks

A quite large number of respondents often consume carbonated soft beverages, 38.13%, which increases the risk of dental caries due to their high content of sugars. As far as the consumption of sweets is concerned, we observe that 58.76% of respondents consume sweets daily, which means a high content of carbohydrates, which, associated with carbonated beverages, favours the carious process.

A total of 167 people surveyed perform their mastication unilaterally and not bilaterally, as normal for a healthy dentition, while preferring at the same time soft food and having bleeding during dental brushing. The same category prefers dental brushing with soft toothbrushes, thus leading to ineffective maintenance.

This demonstrates that these people show signs of dental disease, gingival inflammation and an increased risk of caries formation. Clearly, they are among the patients who fail to visit their dentist frequently and constantly, but only when the pain occurs.

Concerning the consumption of fruit and vegetables, 31% of the respondents rarely consume these foods, which leads to a low intake of vitamins. Vitamin deficiency leads to a decrease in periodontal tissue resistance, from which we conclude that there is a risk of developing gingivitis or various forms of periodontitis.

At the European level, the oro-dental health situation of the population in our country is not optimistic at all. Thus, the report of the European Platform forBetter Oral Health2012 shows that Romania is ranked last in Europe on the budget allocated to oral health and the International Agency for Research on Cancer in 2016 places Romania on the 9th place among the European countries regarding the new cases of oral cancers and on the 8th place as the mortality rate from them. According to the Eurobarometer surveyon the oral health of citizens of EU member states, 41% of Europeans have natural dentition while in Romania only 30% do not have dentures. Most Europeans say they have not experienced dental discomfort in the past year and 15% have had difficulty masticating, unlike in Romania where the percentage is 32%.

A percentage of 7% of European citizens are embarrassed by their dentition, while in Romania the percentage is 16%, which ranks us first in Europe.

While European citizens are visiting a dental practitioner for a routine examination twice a year, on average, in our country about 42.09% do not visit their dentist at all or get an exam once a year at most. Only 20% of patients in the European countries visit the dental practitioner in case of emergency, while in Romania their percentage is 40%.

3. Conclusions and discussions

From this analysis we can draw two directions, namely:

1. Most of the respondents who have had root fillings following carious processes, also declare a frequent consumption of sweets and carbonated beverages, plus the preference for eating soft foods, which prevents the self-cleaning of the teeth. Hence, we conclude that there is a high risk of developing dental caries for them.
2. Lack of interest in dental hygiene and aesthetic appearance. Although a large number of respondents said they were brushing their teeth correctly, twice a day, there are enough situations when tooth brushing is performed once a day or less. Also, a significant number of subjects do not know the additional means of hygiene (mouthwash, dental floss, oral irrigation, etc.). Specialized

information is usually obtained from advertising materials and internet sources that do not always present the most accurate data. The frequency of visiting specialist doctor is quite rare and usually only when the pain occurs and not for preventive purposes.

In an effort to reduce the health gap of the Romanian population compared to the other EU countries, the National Centre for Health Assessment and Promotion, the National Institute of Public Health, the Ministry of Health and the World Health Organization have developed a series of campaigns to promote oral health. First of all, on the occasion of the World Oral Health Day on March 20th, various posters and brochures about the individual behaviour towards their own dental health were printed and distributed. During these campaigns, some information was provided on the proper technique for dental brushing. A set of educational materials that can be used by teachers in the classroom was developed. Some posters to inform children about good dental care practices were similarly designed for schools. There were printed leaflets addressed to pregnant women because it is well known that special attention should be paid to dental hygiene during pregnancy, for the cariogenicity is increased due to the decrease of mineral resources of the body. In addition, during this period, gingivitis often occurs due to hormonal changes. Therefore, the leaflets provide information to the future mother also on the care of the teeth in the newborn. Because they have great importance, these campaigns must continue and address all age groups.

Bibliography

1. Barnes, A. **Personal and Community Health**, Third Edition, Ed. Baillière Tindall, London, 1991.
2. Beaglehole, R., Bonita, R. & Kjellström, T. **Bazele epidemiologiei (trad. Marcu, I.),** Ed. All, Bucharest, 1997
3. Cuculescu, M. **Preventie primara in carie si parodontopatie,** Ed. Didactica si Pedagogica, Bucharest, 2010.
4. Dumitriu, H. T. **Parodontologie**, Ed. Viata medicala romaneasca, Bucharest, 1997.
5. Georgescu, T. **Parodontologie**, Ed. Universitatii din Pitesti, 2004.
6. Iliescu, A. & Gafar, M. **Cariologie si odontoterapie restauratoare**, Ed. Medicala, Bucharest, 2003.
7. Lupan, I, Stepco, E. & Şevcenco, N. **Preventia afectiunilor stomatologice**, Centrul Editorial Poligrafic Medicina, Kishinev, 2014.
8. Miller, D.F. **Dimensions of Comunity Health**, second edition, WM.C.Brown Publishers, USA, 1988.
9. Severineanu, V. **Parodontologie clinica si terapeutica,** Ed. Academiei, Bucharest, 1994.
10. * * * National Institute of Public Health, www.insp.ro, Report, 2012.
11. * * * National Institute of Statistics, www.insse.ro, 2018
12. * * * World Health Organization, www.who.int Oral Health - World Congress on Preventive Dentistry.